Teorema Central del Límite

- TCL: indica que, en condiciones muy generales, la distribución de la suma de v.a. tiende a una distribución normal cuando la cantidad de variables es muy grande.

- Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces, si n grande, la v.a.:

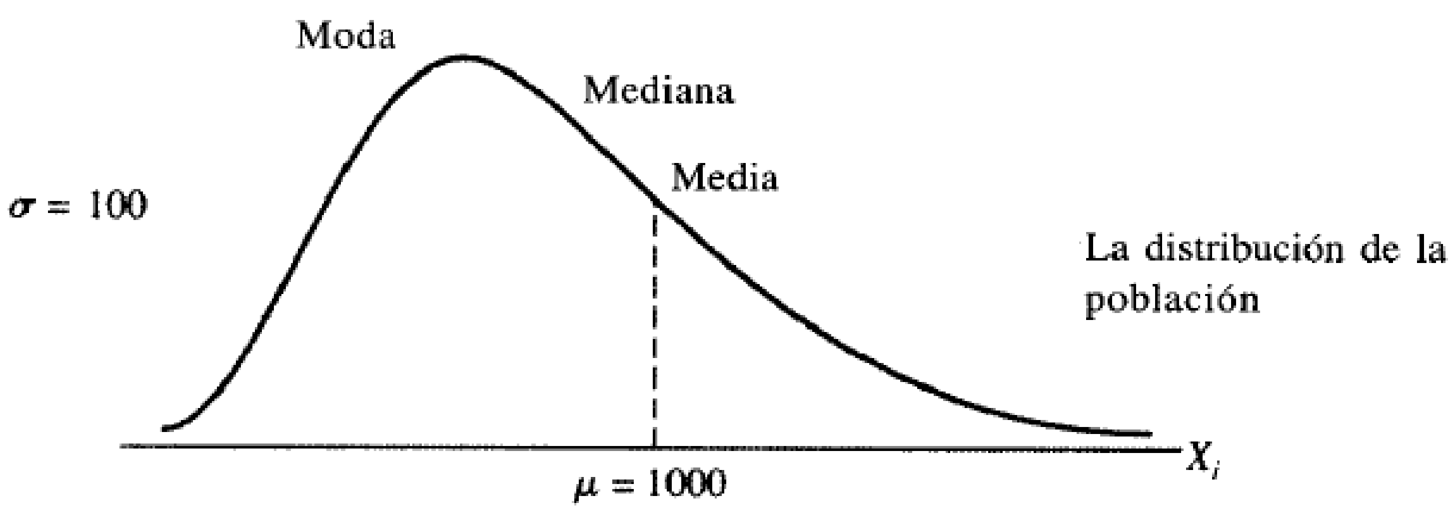
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

tiene aproximadamente una distribución normal con $\mu_{\bar{X}} = \mu$ y $\sigma^2_{\bar{X}} = \sigma^2/n$

Teorema Central del Límite(2)

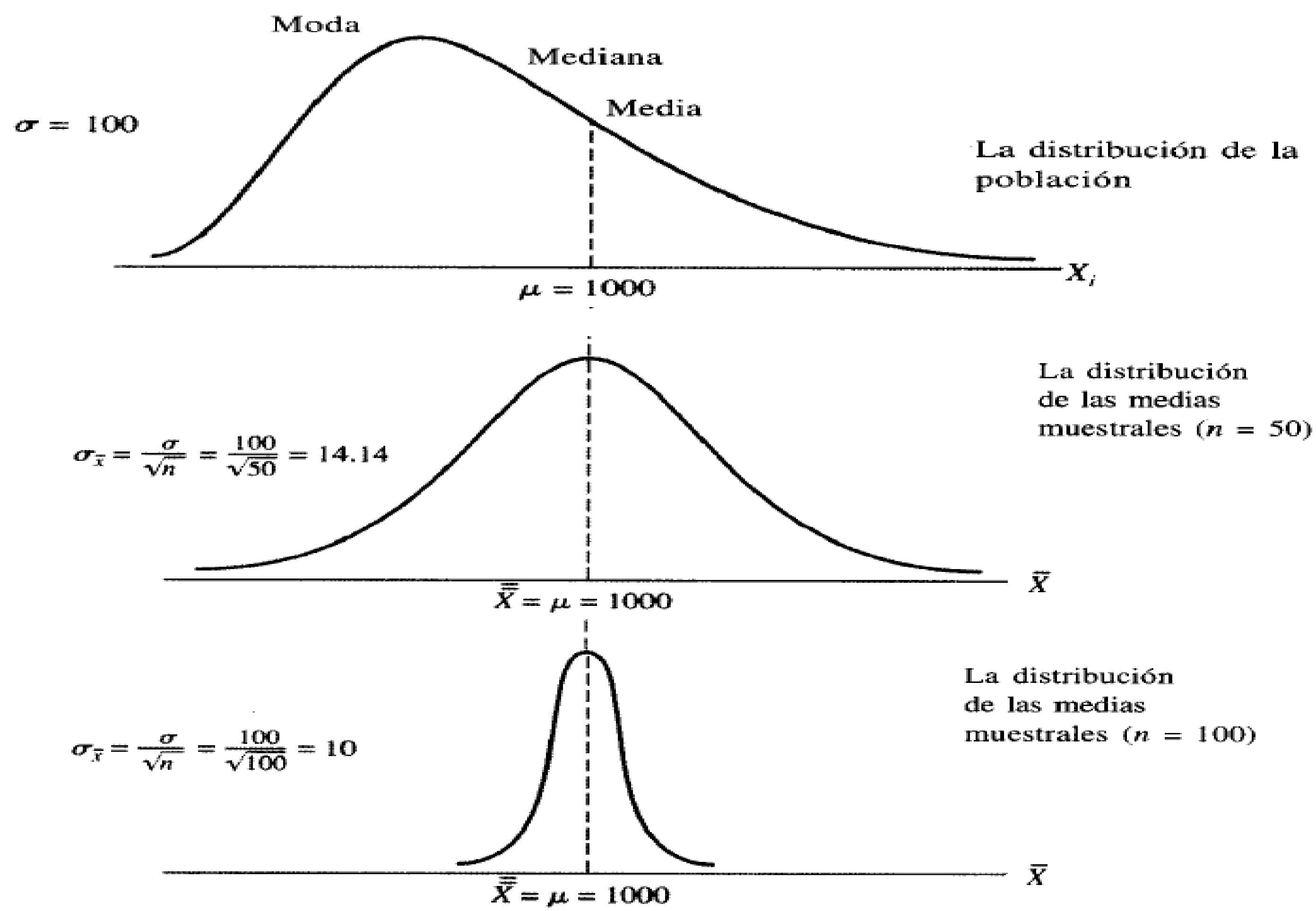
- En particular esto se cumple para la distribución de medias muestrales, aún cuando la distribución original no es normal

- Ej:



Teorema Central del Límite(2)

- Ej:



Teorema Central del Límite(3)

- El TCL permite el uso de la transformación Z para distribuciones muestrales:

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- Esta fórmula es válida cuando la desviación estándar **poblacional** es conocida.

Intervalos de Confianza

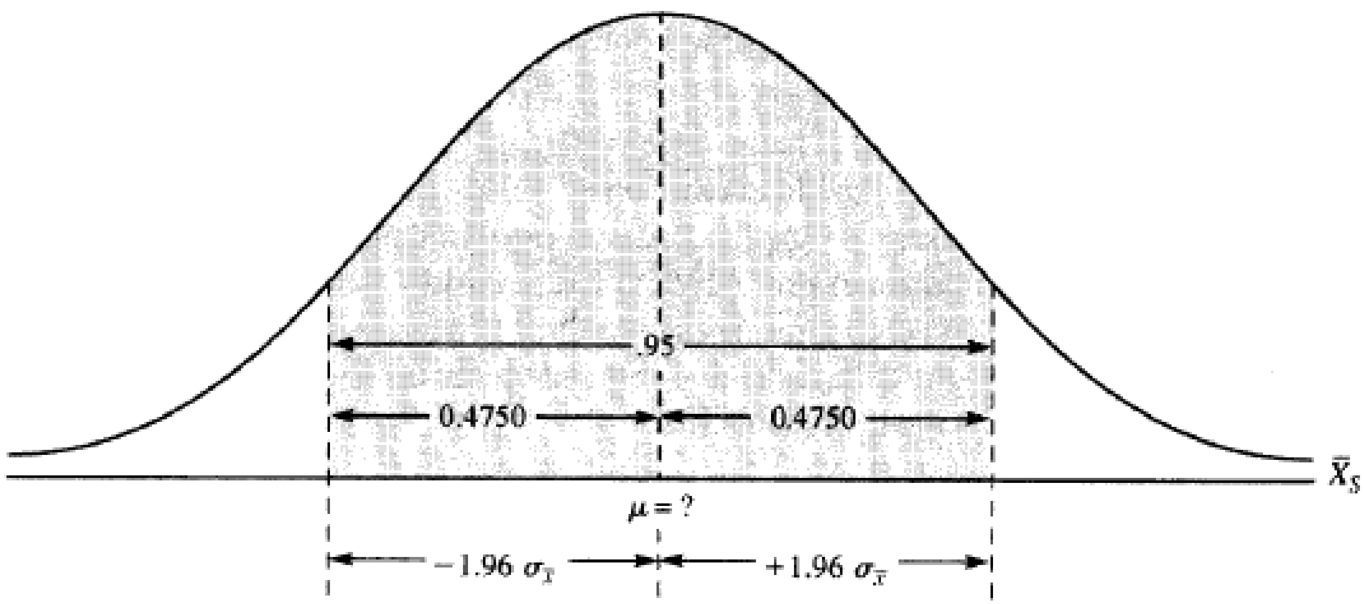
- Un intervalo de confianza denota un rango dentro del cual puede encontrarse un parámetro, y el nivel de confianza que el intervalo considera del parámetro.

- Los niveles de confianza más usados son 90%, 95%, 99%. Se denomina α a la probabilidad de que el intervalo NO contenga el parámetro desconocido.

- Si se desea construir un intervalo para la media muestral de 95%, se debe dividir el área en 2: 47.5% y luego hallar el valor de Z para esta área: Z=1.96.

Intervalos de Confianza(2)

- Este valor representa el número de desviaciones estándar que se permite para la variación de la media muestral, tanto por encima como por debajo.



Intervalos de Confianza – Muestras Grandes*

- Si σ^2 es conocido, el intervalo de confianza para la media muestral μ es $\bar{X} \pm Z\sigma_{\bar{x}}$

- Z=1.96 para un 95% de confianza

- Si σ^2 es desconocido, se utiliza la varianza muestral s^2

$$\text{I.C. para estimar } \mu = \bar{X} \pm Z s_{\bar{X}}$$

*>30 obs.

Intervalos de Confianza – Muestras pequeñas

- El TCL asegura normalidad sólo en muestras grandes

- Cuando las muestras son pequeñas se utiliza la distribución t de Student. Se deben cumplir 3 condiciones:
  - Muestra pequeña
  - σ^2 desconocido (si es conocido se usa Z)
  - Población normal o casi normal (si no se cumple se deben considerar tests no parámetricos)

Intervalos de Confianza – Muestras pequeñas(2)

- Al igual que la distribución Z, t tiene media cero y es simétrica c/r a la media.

- Sin embargo, mientras Z tiene varianza 1, la varianza de t es mayor y depende de la muestra:

$$\sigma^2 = \frac{n-1}{n-3}$$

Intervalos de Confianza – Muestras pequeñas(3)

- El estadístico t se calcula de manera similar a Z:

$$t = \frac{\bar{X} - \mu}{s_{\bar{x}}}$$

- El intervalo de la media muestral de confianza en muestras pequeñas:

$$\text{I.C. para estimar } \mu = \bar{X} \pm (t)(s_{\bar{x}}) = \bar{X} \pm t\frac{s}{\sqrt{n}}$$

Intervalos de Confianza – Muestras pequeñas(4)

- Ej (Wall Street Journal): Una empresa de construcción fue culpada de inflar los comprobantes que registra para los contratos de construcción.
- El contrato estableció que un cierto tipo de trabajo debería promediar US$1150.
- Los directivos de sólo 12 agencias de gobierno dieron testimonio, llegando a una media de US$1275 y desv. est. de US$235. Asumiendo montos normales, ¿un i. de c. de 95% apoyaría el caso legal de la empresa?

Intervalos de Confianza – Muestras pequeñas(5)

- Sol: a un nivel de 95% con 12-1=11 g.l., el valor de t es 2.201

$$\text{I.C. para estimar } \mu = \bar{X} \pm t\,\frac{s}{\sqrt{n}}$$

$$= 1275 \pm (2.201)\,\frac{235}{\sqrt{12}}$$

$$= 1275 \pm 149.31$$

$$\text{US\$}1{,}125.69 \leq \mu \leq \text{US\$}1{,}424.31$$

- Como US$1150 está contenido en el intervalo, se fortalece la defensa de la empresa. Notar que el intervalo es más amplio que si se utiliza la normal.

Intervalos de Confianza – Proporción poblacional

- Las decisiones dependen con frecuencia de parámetros que son binarios (ej proporción de clientes que paga sus créditos)

- La distribución de las prop. muestrales se distribuye normal en muestras grandes (*np, n(1-p)*>5) con media igual p y desv. estandar:

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

- El I.C. para estimar $\pi = p \pm Z s_p$

Intervalos de Confianza – Proporción poblacional

- Ej: el gerente de una estación de televisión debe determinar qué porcentaje de casas tiene más de un televisor.

- Una muestra aleatoria de 500 casa revela que 275 tienen dos o + tvs. ¿Cuál es el i.c. del 90% para estimar la proporción de casas con dos o + tvs?

- Dados los datos p=275/500=0.55 y $s_p = \sqrt{\dfrac{(0.55)(0.45)}{500}} = 0.022$

Intervalos de Confianza – Proporción poblacional

- Ej (cont): de la tabla Z=1.65:

$$\text{I.C. para estimar } \pi = 0.55 \pm (1.65)(0.022)$$
$$= 0.55 \pm 0.036$$
$$0.514 \leq \pi \leq 0.586$$

- El gerente puede tener un 90% de confianza que entre el 51.4% y el 58.6% de las casas de la ciudad tienen más de un televisor.