UNIVERSIDAD
NACIONAL
DE COLOMBIA

Estadística I

Inferencia Estadística

El problema de la inferencia estadística surge cuando se desea hacer generalizaciones acerca de una población y se dispone solo de una muestra. Recuerde que la inferencia estadística permite obtener conclusiones acerca de parámetros poblacionales a partir del análisis de datos. Cualquier inferencia acerca de un parámetro poblacional involucra cierto grado de incertidumbre ya que está basado en una porción de la población (muestra). En estadística los dos tipos de inferencia más comunes son: Estimación por intervalos y Pruebas de hipótesis. La estimación de parámetros por intervalos tiene como objetivo principal obtener una estimación de un valor desconocido con un buen grado de precisión.

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

La estimación por intervalos de un parámetro poblacional consiste en el cálculo de 2 valores que componen los extremos de un intervalo que debe contener al parámetro con una alta probabilidad. Estos extremos dependen usualmente de la distribución del estimador y de los datos. Una característica deseable en un estimador por intervalo es que contenga el parámetro con una confianza pre-especificada (que casi nunca será del 100% debido a la variabilidad presente en la muestra).

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

$V[\bar{X}] = \sigma^2/n$

μ

$\bar{X} = \frac{\sum x_i}{n}$ T1.

TLC: $n \geq 30$ $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{aprox}{\sim} N(0,1)$

Sea θ un parámetro de interés y $\hat{\theta}$ un estimador puntual de θ. Un estimador por intervalos es de la forma (l, u), donde l y u dependen de $\hat{\theta}$ y su distribución.

Cada muestra aleatoria proporcionará valores distintos para $\hat{\theta}$ y en consecuencia para l y u, luego los extremos del intervalo en cuestión se convierten en variables aleatorias, las cuales denotaremos como L y U. El intervalo (L, U) es llamado **Intervalo Aleatorio.**

$\bar{X}$

$(2, 5)$

$\boxed{\bar{x} = 21.}$

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

Usando $\hat{\theta}$ y su distribución es posible determinar L y U tales que:

$$P(L < \theta < U) = 1 - \alpha, \quad \alpha \in (0,1),\ \alpha \text{ dado}$$

Para una muestra en particular, se obtiene el intervalo $(\overset{\#}{l}, \overset{\#}{u})$ donde se espera esté el verdadero valor del parámetro θ. El intervalo (l, u) se conoce como **intervalo de confianza** al $100(1-\alpha)\,\%$ para θ. l y u son llamado límites de confianza.

Intervalos de Confianza

Obtención de un I.C:

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución $f(x)$ que depende de un parámetro desconocido θ. Sea $\hat{\theta}$ un estimador puntual para θ, así que $\hat{\theta}$ es una función de la m.a, digamos $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$. Suponga además que la función de $\hat{\theta}$ puede depender de θ. Entonces, para $\boxed{\alpha} \in (0,1)$ dado se pueden encontrar las constantes a y b tal que:

$$P\left(a < h(X_1, X_2, \ldots, X_n) < b\right) = 1 - \alpha$$

Suponga que de esta expresión se puede despejar a θ, de esta manera se obtiene:

$$P\left(L(X_1, X_2, \ldots, X_n) < \theta < U(X_1, X_2, \ldots, X_n)\right) = 1 - \alpha$$

Para una muestra particular se calcula $l = l(X_1, X_2, \ldots, X_n)$ y $u = u(X_1, X_2, \ldots, X_n)$; entonces el intervalo (l, u) es un intervalo de confianza al $100(1-\alpha)\,\%$ para θ.

Intervalos de Confianza

Nivel de Confianza: $(1 - \alpha)$

- Un 95% de confianza implica que el 95% de todas las muestras daría un intervalo que incluye el parámetro que se está estimando y que sólo el 5% de las muestras darían un intervalo erróneo.

- Los niveles de confianza que se usan con más frecuencia son 90%, 95%, 98% y 99%.

- Mientras más alto es el nivel de confianza, más fuerte es la creencia de que el valor del parámetro que se está estimando queda dentro del intervalo.

Intervalos de Confianza

Precisión ε :

- El ancho del intervalo da información sobre la precisión de la estimación por intervalos.

- Entre mayor sea la longitud de un I.C menor será su precisión, lo que nos indica que existe una gran cantidad de incertidumbre sobre el valor que se está estimando.

- Entre menor sea la longitud del I.C mayor será su precisión. Lo ideal es tener intervalos de confianza angostos con una alta confianza.

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

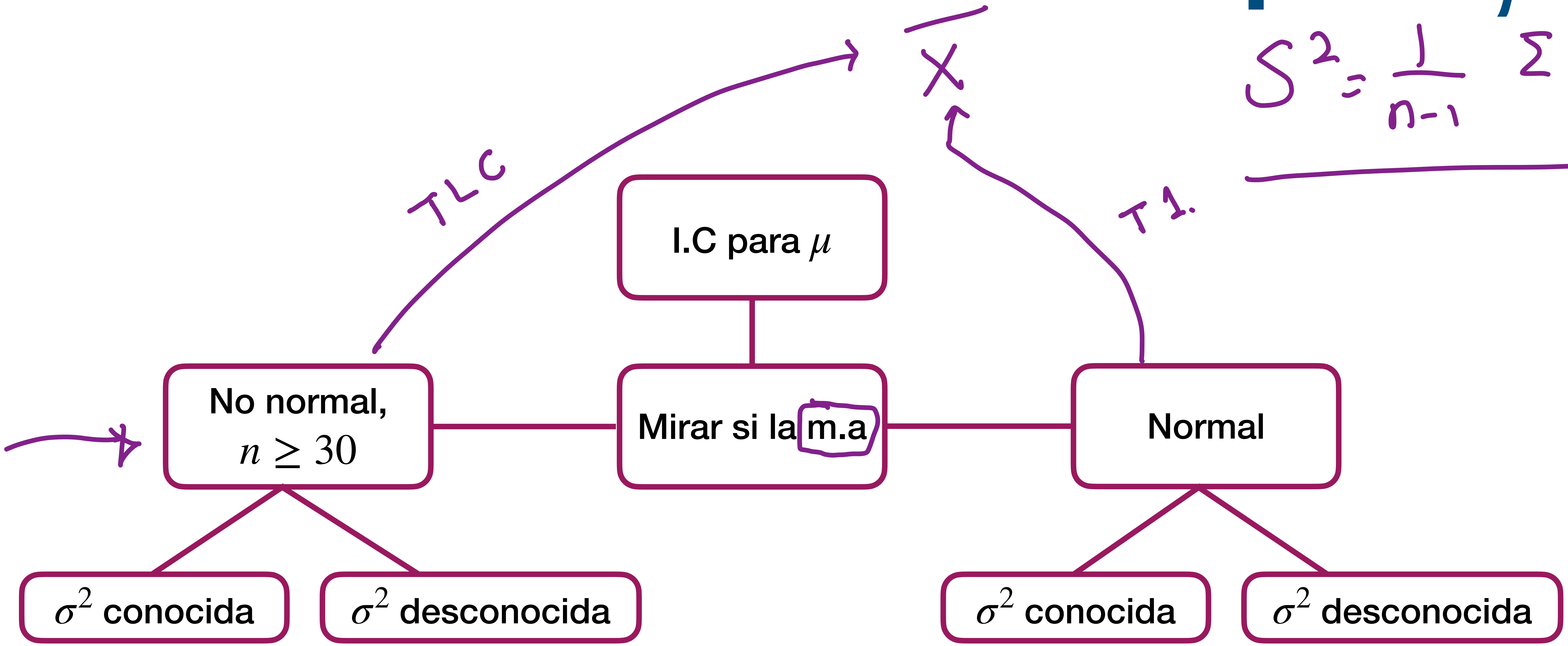
Precisión

Note que en un intervalo de confianza, el nivel de confianza y la precisión tienen una relación inversa. Lo que implica que, mientras más alto es el grado de confianza, mas ancho es el intervalo, mientras que un intervalo preciso puede acarrear una confiabilidad baja.

Una estrategia atractiva es especificar tanto el nivel de confianza deseado como el ancho del intervalo y luego determinar el tamaño de la muestra necesario.

Intervalos de Confianza para μ

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$



Intervalos de Confianza para μ

$n \geq 30$

Caso I: m.a de una población no normal con n grande

☑ σ^2 **conocida:**

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con media μ y varianza σ^2 conocida. El interés recae en construir un intervalo al $100(1-\alpha)\%$ para μ. Debemos hallar L y U tal que:

$$P(L < \underline{\mu} < U) = \underbrace{1-\alpha}_{\text{fijo}}$$

Intervalos de Confianza para μ

Caso I: m.a de una población no normal con n grande

☑ σ^2 **conocida:**

$$1-\alpha = P(L<\mu<U) = P(-L>-\mu>-U)$$

$$= P(-U<-\mu<-L) = (\bar{X}-U<\bar{X}-\mu<\bar{X}-L)$$

$$= P\left(\frac{\bar{X}-U}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-L}{\sigma/\sqrt{n}}\right) = P\left(Z_1<Z<Z_2\right)$$

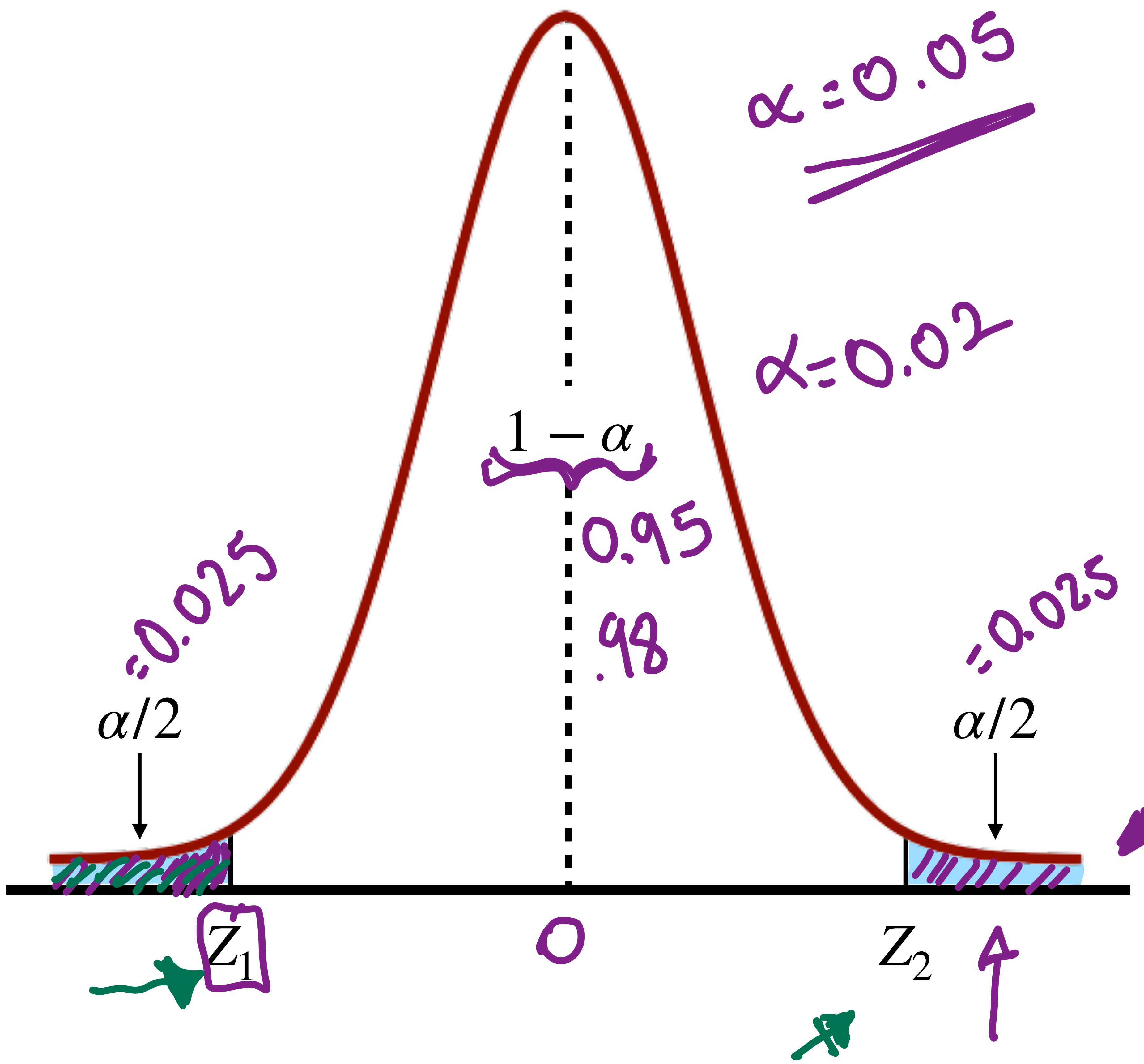
Donde $Z_1 = \dfrac{\bar{X}-U}{\sigma/\sqrt{n}}$ y $Z_2 = \dfrac{\bar{X}-L}{\sigma/\sqrt{n}}$

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza para μ

Caso I: m.a de una población no normal con n grande

☑ σ^2 **conocida:**



$\alpha = 0.05$

$\alpha = 0.02$

Note que $P(Z < Z_1) = P(Z > Z_2) = \alpha/2$

Como se tiene una distribución normal:

$Z_1 = -Z_{\alpha/2}$ y $Z_2 = Z_{\alpha/2}$

Luego:

$$\frac{\bar{X} - \boxed{U}}{\sigma/\sqrt{n}} = -Z_{\alpha/2} \rightarrow \boxed{U = \bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

$Z_{0.025} = 1.96$

$-Z_{0.025} = -1.96$

$P(Z < Z_1) = \alpha/2 \rightarrow 0.025.$

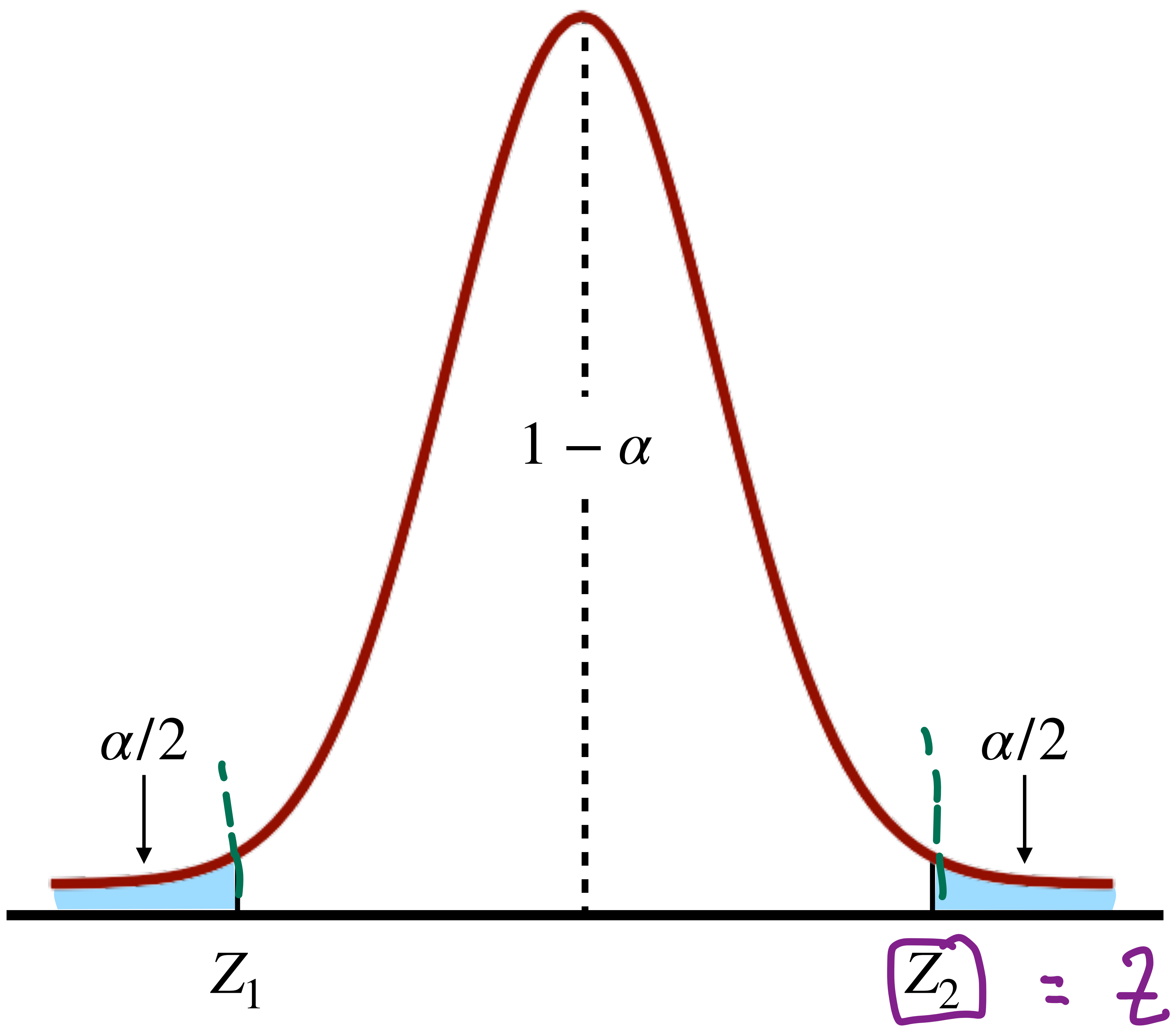
$P(Z > Z_2) = \boxed{0.025}$

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza para μ

Caso I: m.a de una población no normal con n grande

☑ σ^2 **conocida:**



$= z_{\alpha/2}$

$$\frac{\bar{X}-L}{\sigma/\sqrt{n}} = Z_{\alpha/2} \rightarrow \boxed{L = \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

Así para una muestra particular, se calculan los valores de L y U, de esta manera, un I.C aproximado al $100(1-\alpha)\%$ para μ es:

$$\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$\boxed{\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

Intervalos de Confianza para μ

Caso I: m.a de una población no normal con n grande

$\sigma^2 \rightarrow S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

σ^2 desconocida:

Cuando no se conoce σ^2 se usa la varianza muestral S^2. Así, un I.C aproximado al $100(1-\alpha)\,\%$ para μ es:

$$\bar{X} \pm Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

Intervalo de confianza para P

Un I.C aproximado al $100(1-\alpha)\,\%$ para la probabilidad de éxito P de una binomial está dada por:

x: # éxitos en n ensayos

$$\boxed{\hat{P} \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \text{ donde } \hat{P} = \frac{X}{n}}$$

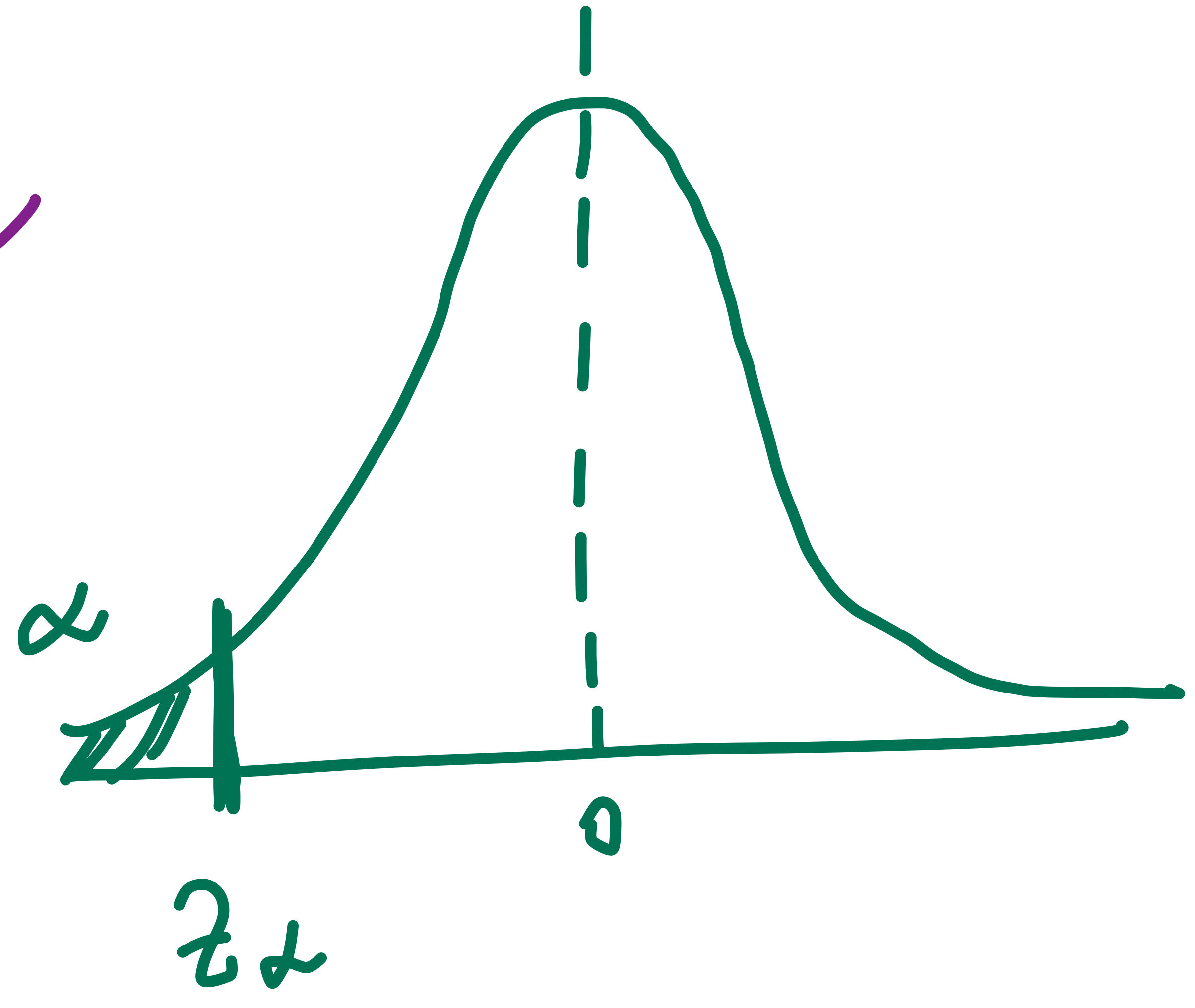
$\frac{x}{n} = \hat{p}$

Intervalos de Confianza para μ

I.C unilaterales:

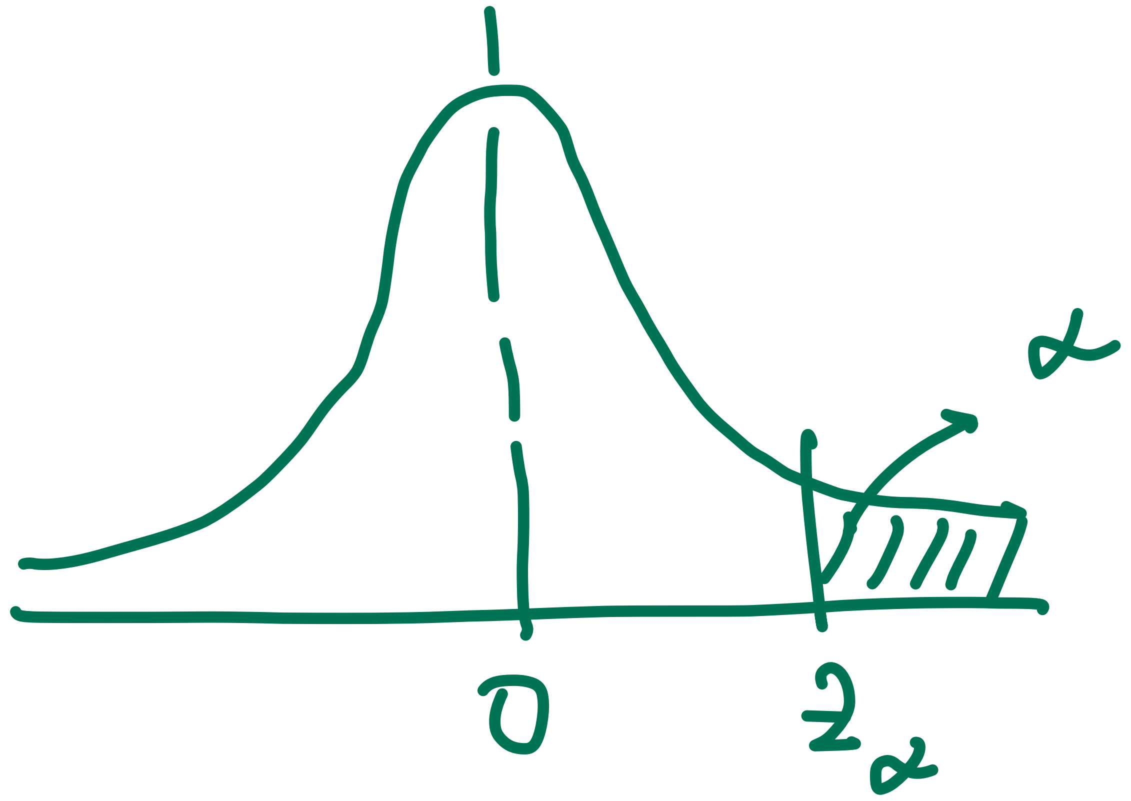
Un I.C unilateral inferior al $100(1-\alpha)\,\%$ para μ es:

$$\left(\bar{X} - \underbrace{Z_\alpha \frac{S}{\sqrt{n}}}, + \infty\right)$$



Un I.C unilateral superior al $100(1-\alpha)\,\%$ para μ es:

$$\left(-\infty, \bar{X} + \underbrace{Z_\alpha \frac{S}{\sqrt{n}}}\right)$$



Intervalos de Confianza para μ

I.C para μ

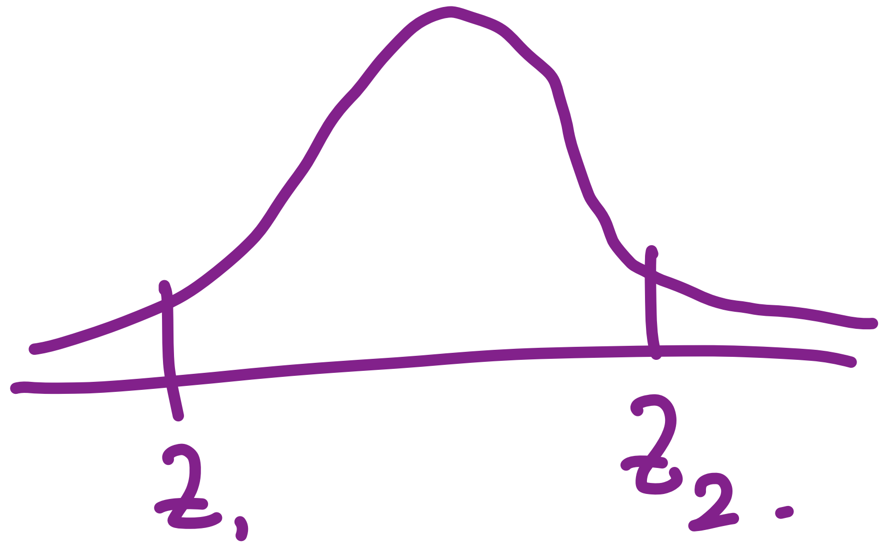
Mirar si la m.a

- No normal, $n \geq 30$
  - σ^2 conocida

    ① $\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
  - σ^2 desconocida

    ② $\bar{X} \pm Z_{\alpha/2}\frac{S}{\sqrt{n}}$
- Normal
  - σ^2 conocida
  - σ^2 desconocida

$\hat{P} \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$



$z_{\alpha/2}$

$$\text{Ancho} = \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$= \cancel{\bar{x}} + z_{\alpha/2}\,\sigma/\sqrt{n} - \cancel{\bar{x}} + z_{\alpha/2}\,\sigma/\sqrt{n}$$

$$= 2\, z_{\alpha/2}\,\sigma/\sqrt{n}$$

Intervalos de Confianza

Precisión: $\varepsilon = \dfrac{\text{Ancho}}{2} = z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

☑ **Tamaño de muestra:**

Ahora suponga que se desea saber o estimar el tamaño de la muestra necesario para que con una confianza $1-\alpha$ la media muestral se encuentre a ε unidades alrededor de la media de la población μ.

Es decir, se quiere que:

$$\boxed{P\left(\left|\bar{X}-\mu\right| < \boxed{\varepsilon}\right) = 1-\alpha}$$

$$P\left(-\varepsilon < \bar{X}-\mu < \varepsilon\right)$$

Intervalos de Confianza

Tamaño de muestra:

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

$$P\left(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = P\left(\left|\bar{X} - \mu\right| < \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\varepsilon}\right)$$

Así que

$$\boxed{\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}} \rightarrow \boxed{n = \left[\frac{Z_{\alpha/2}\sigma}{\varepsilon}\right]^2}$$

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

Ejemplo:

Se tomo una muestra de 36 y se obtuvo un promedio de 2.6 y una desv. estandar de 0.3 $\rightarrow s$

Se encuentra que la concentración promedio de zinc que se obtiene a partir de una muestra de mediciones de zinc en 36 sitios distintos de un río es de 2.6 gramos por milímetros.

a. Encuentre los I.C aproximados del 95% y 99% para la concentración media de zinc en el río. Suponga que la desviación estándar de la población es 0.3 $\rightarrow \sigma$

b. ¿Qué tan grande debe ser el tamaño de muestra si queremos tener un 95% de confianza de que nuestra estimación de μ difiera por menos de 0.05.

Intervalos de Confianza

a) $\bar{x} = 2.6$, $n = 36$, $\sigma = 0.3$

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

• 95% | $1 - \alpha = 0.95$

$\alpha = 0.05$ | $z_{\alpha/2} = z_{0.025} = 1.96$

$\alpha/2 = 0.025$

Así, el I.C del 95% está dado por

$$2.6 \pm 1.96 \left(\frac{0.3}{\sqrt{36}}\right)$$

$(2.5025, 2.6975)$

• 99% | $1 - \alpha = 0.99$

$\alpha = 0.01$ $\quad z_{0.005} = 2.57$

$\alpha/2 = 0.005$

Así, el I.C del 99% está dado por:

$$2.6 \pm 2.57 \left(\frac{0.3}{\sqrt{36}}\right)$$

$(2.4715, 2.7285)$

Intervalos de Confianza

$\varepsilon = 0.05$ $\qquad 1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$Z_{\alpha/2} = 1.96$

$$n = \left[\frac{Z_{\alpha/2}\,\sigma}{\varepsilon}\right]^2$$

$$n = \left[\frac{1.96 * 0.3}{0.05}\right]^2 = 138.29 \approx 139$$

Intervalos de Confianza

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma \approx S}{\sqrt{n}}$$

Ejemplo:

El número de abolladuras en láminas de aluminio sigue una distribución Poisson. De una muestra aleatoria de 100 láminas, se obtiene un promedio de 5 abolladuras. Calcule un intervalo de confianza aproximado del 98% para el promedio de abolladuras real.

X: # abolladuras en una lámina $\quad X \sim P(\lambda)$

$\bar{X} = 5 = \hat{\lambda}$

$E[X] = \hat{\lambda}$ $\quad S^2 = 5$

$Var[X] = \hat{\lambda}$

$1 - \alpha = 0.98$

$\alpha = 0.02$

$\alpha/2 = 0.01$

$z_{0.01} = 2.33$

$$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}}$$

$$5 \pm 2.33 \frac{\sqrt{5}}{\sqrt{100}} \Rightarrow (4.47, 5.52)$$

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

Intervalos de Confianza

Ejemplo:

Un fabricante de pilas alcalinas sabe que el tiempo de duración, en horas, de las pilas que fabrica tiene una varianza 3600. Con una muestra de su producción, elegida al azar, y un nivel de confianza del 95% ha obtenido para la media el intervalo de confianza (372.6, 392.2).

a. Calcule el valor que obtuvo para la media muestral y el número de pilas alcalinas estudiadas.

b. Se cree que el tiempo de duración de las pilas alcalinas es superior a 375 horas. ¿Qué se puede concluir con un intervalo de confianza al 90%?

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

$1-\alpha = 0.95$, $z_{0.025} = 1.96$

a) $\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$$\varepsilon = \frac{\text{Ancho}}{2} = \frac{392.2 - 372.6}{2} = 9.8$$

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 9.8$$

$$1.96 \frac{\sqrt{3600}}{\sqrt{n}} = \boxed{9.8}$$

$$\frac{60}{\sqrt{n}} = \frac{9.8}{1.96} \Rightarrow \frac{60}{\sqrt{n}} = 5$$

$$\Rightarrow \sqrt{n} = \frac{60}{5}$$

$$n = \left(\frac{60}{5}\right)^2 = 144$$

$$LI + LS = \bar{X} - \cancel{z_{\alpha/2} \frac{\sigma}{\sqrt{n}}} + \bar{X} + \cancel{z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

$$= 2\bar{X}$$

$$\frac{LI + LS}{2} = \bar{X} \checkmark$$

$$\frac{372.6 + 392.2}{2} = 382.4 = \bar{X}$$

Intervalos de Confianza

$$\bar{X} \pm \varepsilon$$

$$\bar{X} + \varepsilon = 392.2$$

$$\bar{X} + 9.8 = 392.2$$

$$\bar{X} = 392.2 - 9.8$$

$$\bar{X} = 382.4$$

b) I.C unilateral para μ:

$$\left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}},\ \infty\right)$$

Afirmación ¿$\mu > 375$?

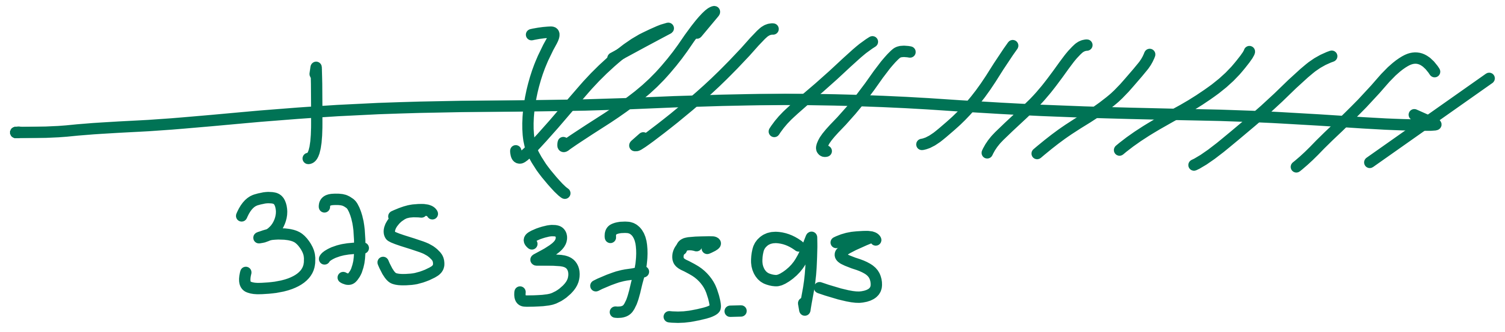
$$1 - \alpha = 0.9$$

$$\alpha = 0.1$$

$$z_{0.1} = 1.28$$

$$\left(382.4 - 1.28 \frac{60}{\sqrt{144}},\ \infty\right)$$

$$(375.95,\ \infty)$$



Intervalos de Confianza

Ejemplo:

Considere 1500 intervalos de confianza del 95% para la pérdida por carga parásita promedio μ de cierto tipo de motor de inducción. Suponga que el conjunto de datos de los cuales están basados los intervalos se selecciona de manera independiente.

(a) ¿Cuántos de estos intervalos se espera que capturen el valor correspondiente de μ?

(b) ¿Cuál es la probabilidad de que entre 1410 y 1450 de estos intervalos inclusive contengan el valor correspondiente de μ?

$n = 1500$

X: # intervalos que contienen el μ real.

$X \sim \text{Binomial}(n = 1500,\ p = \underline{0.95})$

Intervalos de Confianza

a) $E[X] = np = (1500)(0.95) = 1425$

b) $P(1410 \le X \le 1450) \approx P(1409.5 \le X \le 1450.5)$

$$\approx P\left(\frac{1409.5 - 1425}{\sqrt{1425(0.05)}} \le Z \le \frac{1450.5 - 1425}{\sqrt{1425(0.05)}}\right)$$

$$\frac{\# - np}{\sqrt{np(1-p)}}$$

$$\approx P[-1.83 \le Z \le 3.02] = \Phi(3.02) - \Phi(-1.83)$$

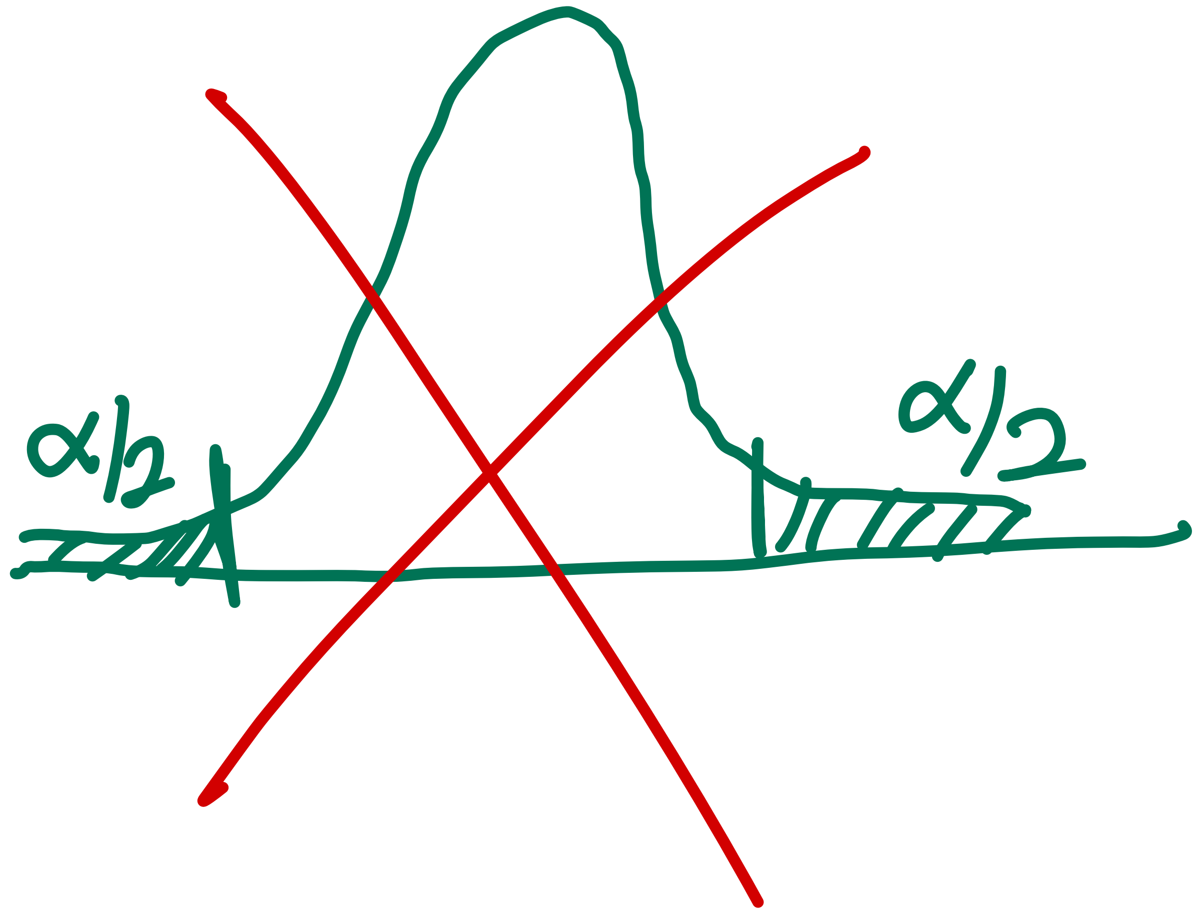
$$= 0.96.$$

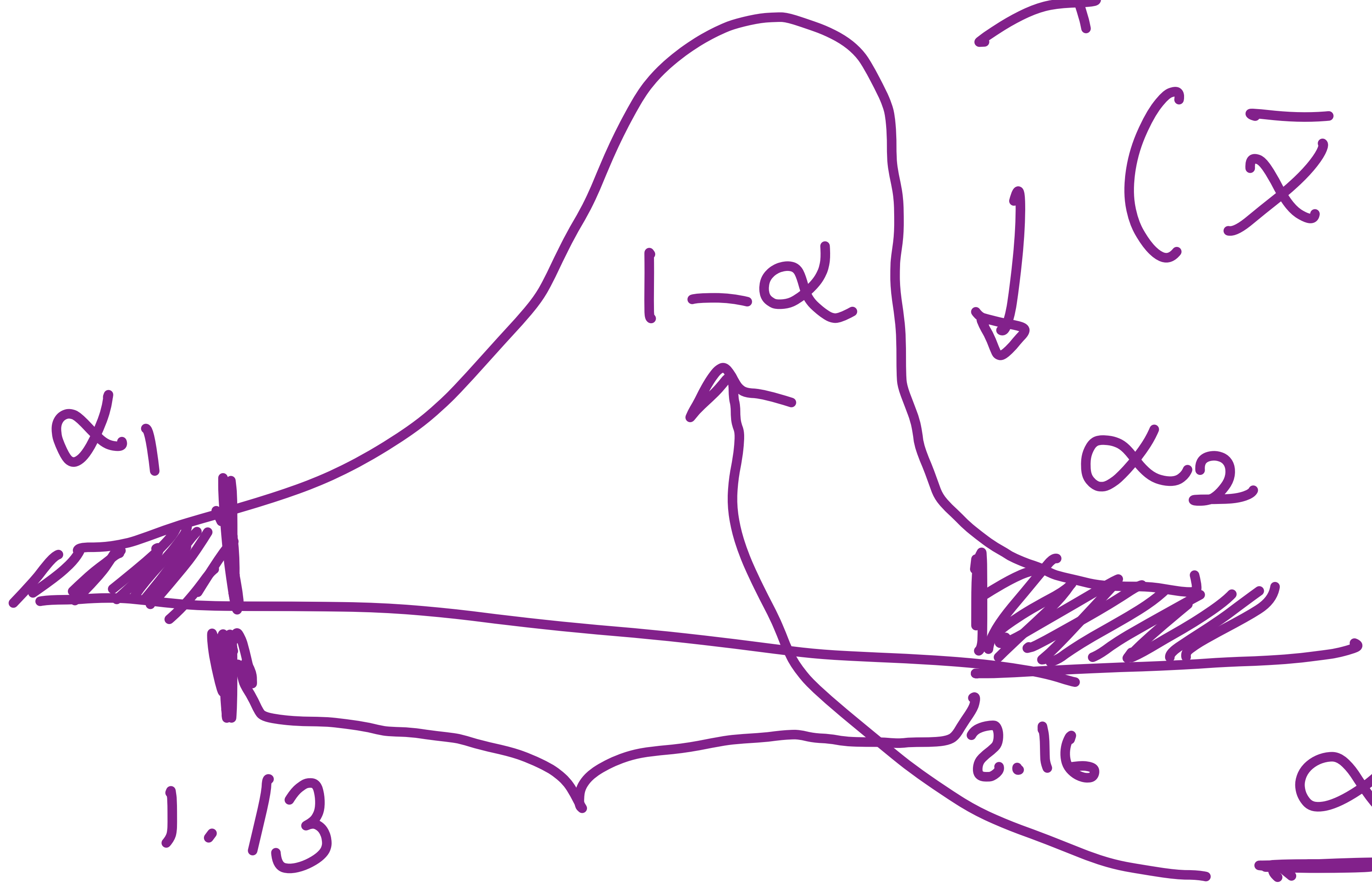
Intervalos de Confianza

Intervalos de Confianza

Suponga que la concentración de cloruro en la sangre tiene una distn normal con $\sigma = 5$ mmol/L. De una muestra de 22 mediciones se obtuvo que la concentración promedio de cloruro en la sangre es de 104 mmol/L. Calcule la probabilidad de cobertura del I.C $\left(\underbrace{\bar{x}-1.2}, \bar{x}+2.3\right)$ ¿y si es $(\bar{x}-1.2, \infty)$?

$(\bar{x}-1.2, \bar{x}+2.3) \longrightarrow$ bilateral asimétrico.





el I.C es de la forma

$$\left(\bar{x} - \underbrace{z_{\alpha_1}\frac{\sigma}{\sqrt{n}}}_{1.2}, \ \bar{x} + \underbrace{z_{\alpha_2}\frac{\sigma}{\sqrt{n}}}_{2.3}\right)$$

$\alpha = \alpha_1 + \alpha_2$

Intervalos de Confianza

Para α_1:

$$1.2 = z_{\alpha_1}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow z_{\alpha_1} = \frac{1.2\sqrt{n}}{\sigma} = \frac{1.2\sqrt{22}}{5} = 1.13$$

$$\alpha_1 = P(Z < -1.13) = 0.12$$

Para α_2:

$$2.3 = z_{\alpha_2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow z_{\alpha_2} = \frac{\sqrt{22} \times 2.3}{5} = 2.16$$

Luego:

$$\alpha_2 = P(Z > 2.16)$$
$$= 0.015$$

Así

$$\alpha = \alpha_1 + \alpha_2 = 0.12 + 0.015$$
$$= 0.135$$

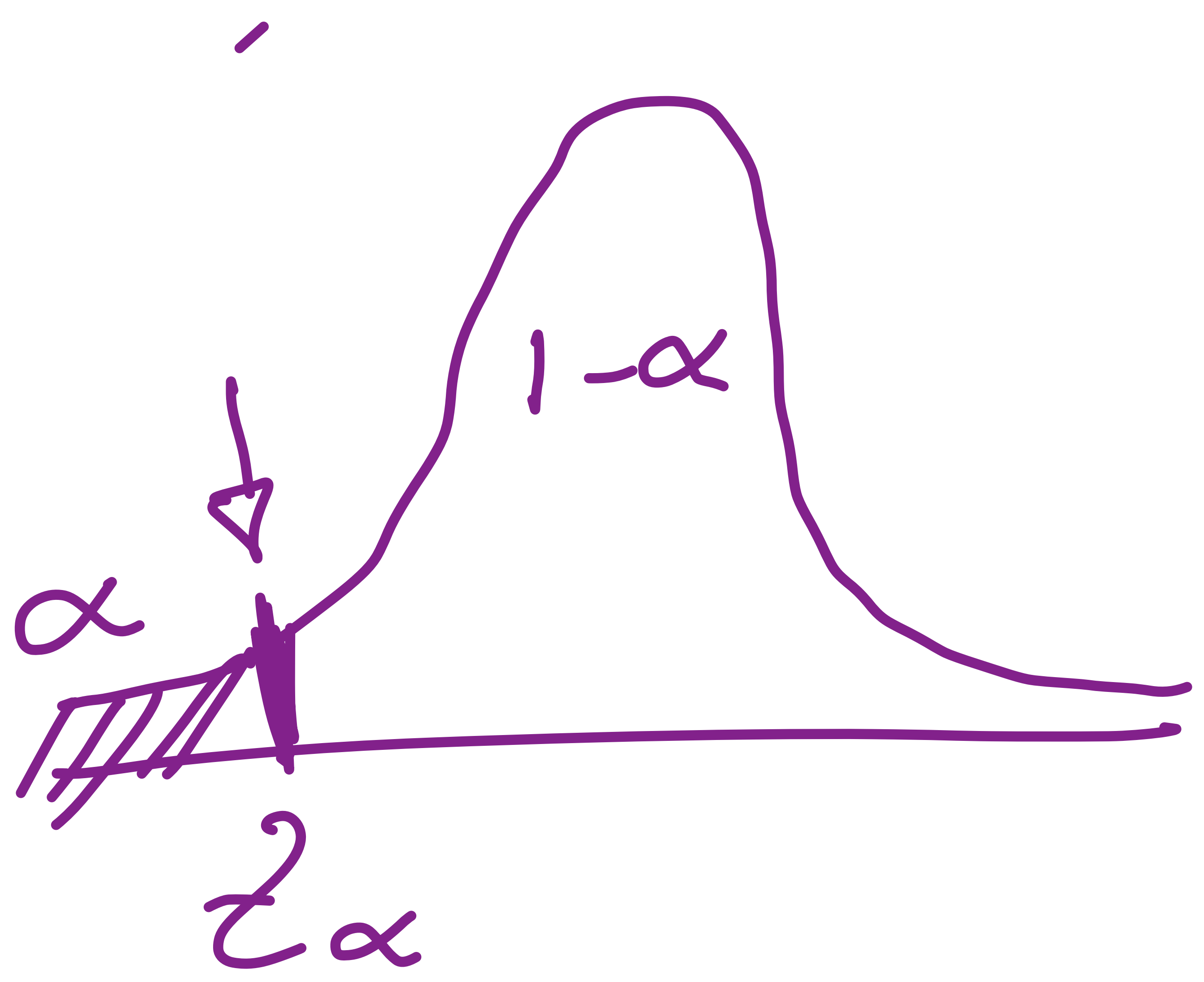
Así, $1-\alpha = 0.86$

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

$$\left(\underbrace{\bar{X} - 1.2}, \infty\right)$$

$$\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$



Gracias

Clase preparada por: Verónica Guarín