Problemas de Estimación de Una y Dos Muestras

UCR – ECCI
CI-0115 Probabilidad y Estadística
Prof. Kryscia Daviana Ramírez Benavides

Inferencia Estadística

- La teoría de la **inferencia estadística** consiste en aquellos métodos por los que se realizan inferencias o generalizaciones acerca de una población.
- La inferencia estadística se puede dividir en dos áreas principales: **estimación** y **prueba de hipótesis**.
- En este capítulo se tratará sobre el área de estimación.

Estimación de la Media
Una Muestra

- Si la muestra se selecciona de una población normal o, a falta de esta, si n es suficientemente grande, se puede establecer un intervalo de confianza para μ al considerar la distribución muestral de $\overline{X}$.
- La estimación de la media de la población (μ), se hace por medio del teorema del límite central.
- De acuerdo a este teorema se puede establecer la distribución muestral de $\overline{X}$, que está distribuida de forma aproximadamente normal con $\mu_{\overline{X}} = \mu$ y desviación estándar $\sigma_{\overline{X}} = \sigma / \sqrt{n}$.

Estimación de la Media
Una Muestra (cont.)

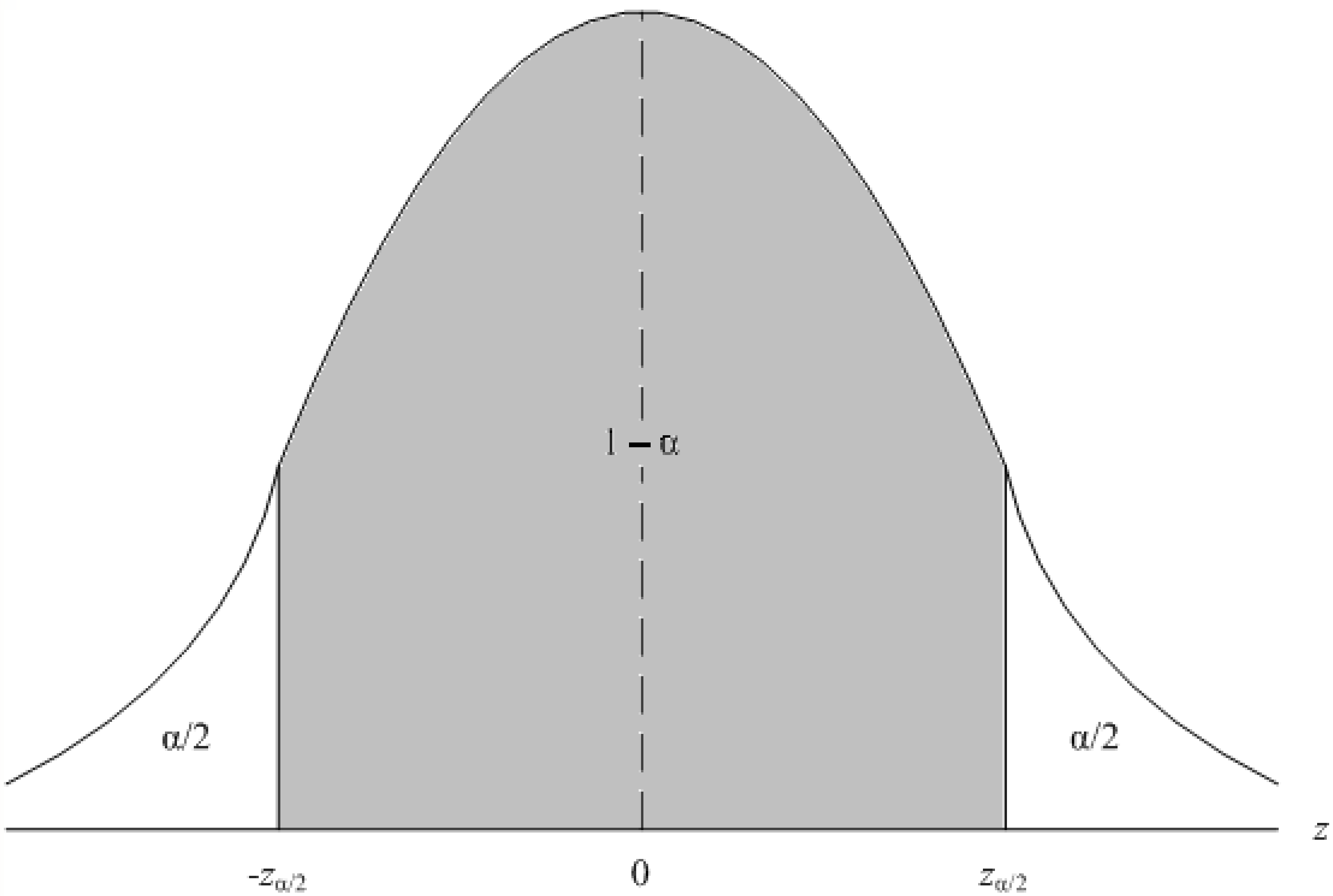
- Al escribir $z_{\alpha/2}$ para el valor z el cual se tiene un área de $\alpha/2$ a la derecha, se puede ver que

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

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

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

Estimación de la Media
Una Muestra (cont.)



Estimación de la Media
Una Muestra (cont.)

- Al multiplicar cada término en la desigualdad por $\sigma/\sqrt{n}$, después restar $\overline{X}$ y multiplicar por -1, se obtiene:

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

- Se selecciona una muestra aleatoria de tamaño *n* de una población cuya varianza σ^2 se conoce y se calcula la media de la muestra para obtener un intervalo de confianza de $(1 - \alpha)100\%$.

Estimación de la Media
Una Muestra (cont.)

- **Intervalo de Confianza de μ con σ conocida.** Si $\bar{x}$ es la media de una muestra aleatoria de tamaño *n* de una población con varianza σ^2, conocida, un intervalo de confianza de $(1 - \alpha)$ 100% para μ está dado por

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

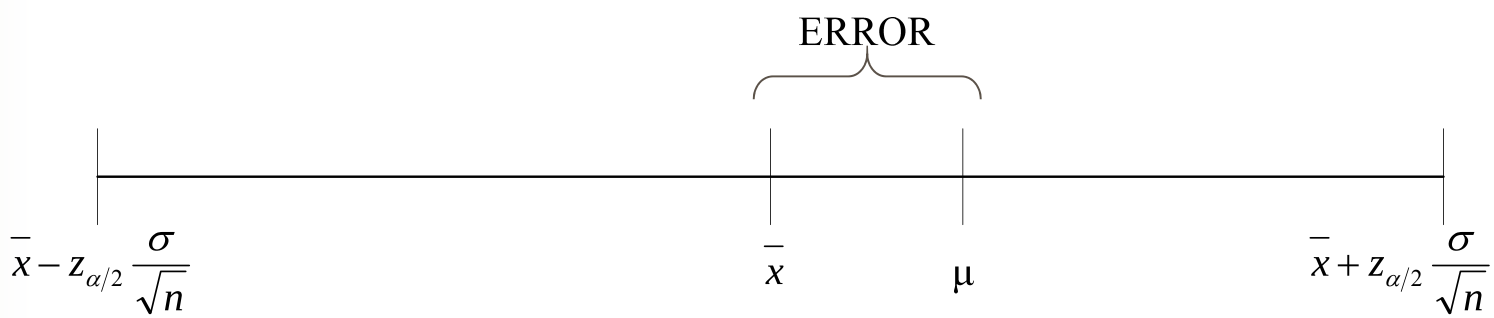
- Donde $z_{\alpha/2}$ es el valor *z* que deja un área de $\alpha/2$ a la derecha.

Estimación de la Media
Una Muestra (cont.)

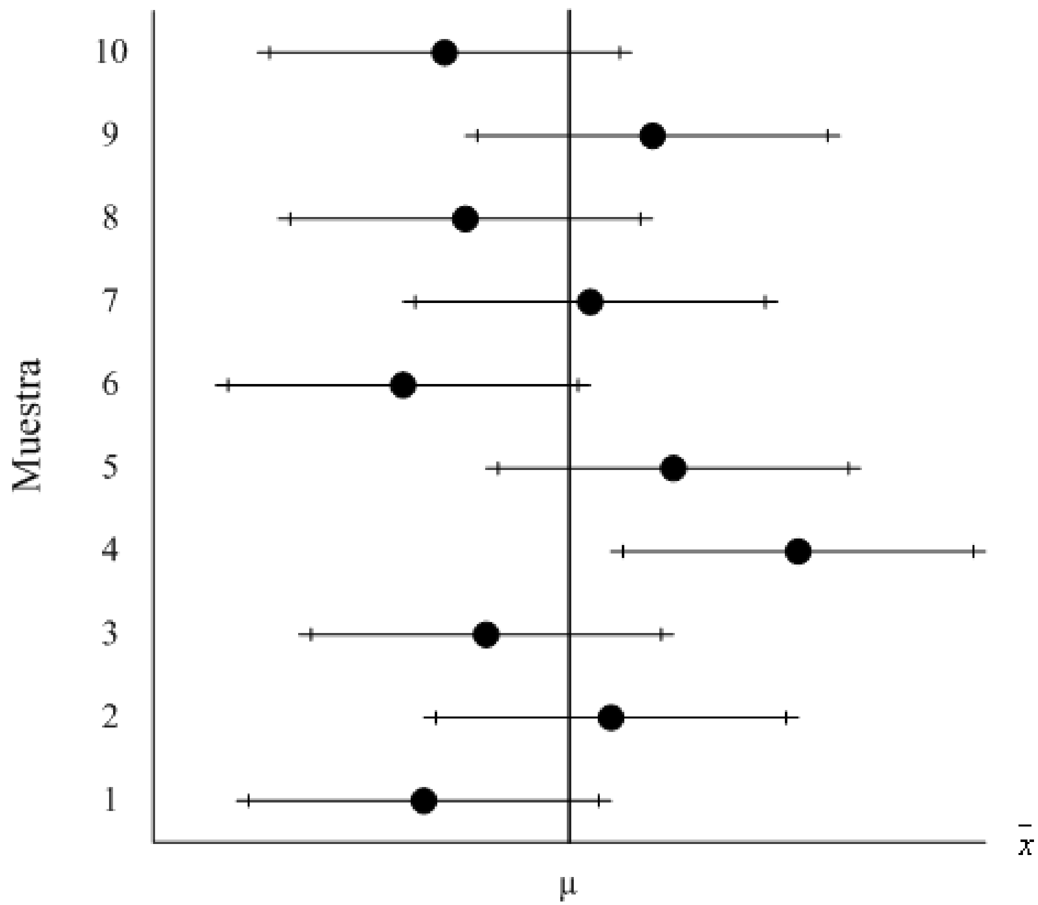
- Para muestras pequeñas que se seleccionan de poblaciones no normales, no se puede esperar que el grado de confianza sea preciso.
- Para muestras de tamaño $n \geq 30$, sin importar la forma de la mayor parte de las poblaciones, la teoría de muestreo garantiza buenos resultados.

Estimación de la Media
Una Muestra (cont.)

- Si μ es realmente el valor central del intervalo, entonces $\bar{x}$ estima μ sin error. La mayor parte de las veces, $\bar{x}$ no será exactamente igual a μ y la estimación puntual es errónea.
- La magnitud de este error será el valor absoluto de la diferencia entre μ y $\bar{x}$, y se puede tener (1 – α)100% de confianza de que esta diferencia no excederá $z_{\alpha/2}\,\sigma/\sqrt{n}$.



Estimación de la Media
Una Muestra (cont.)



Estimación de la Media Una Muestra (cont.)

- **Teorema.** Si se utiliza $\bar{x}$ como una estimación de μ, se puede tener una confianza de (1 – α)100% de que el error no excederá de $z_{\alpha/2}\,\sigma/\sqrt{n}$.
- **Teorema.** Si se utiliza $\bar{x}$ como una estimación de μ, se puede tener una confianza de (1 – α)100% de que el error no excederá una cantidad específica *e* cuando el tamaño de la muestra es

$$n = \left(z_{\alpha/2} \frac{\sigma}{e} \right)^2 \qquad e = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Estimación de la Media Una Muestra (cont.)

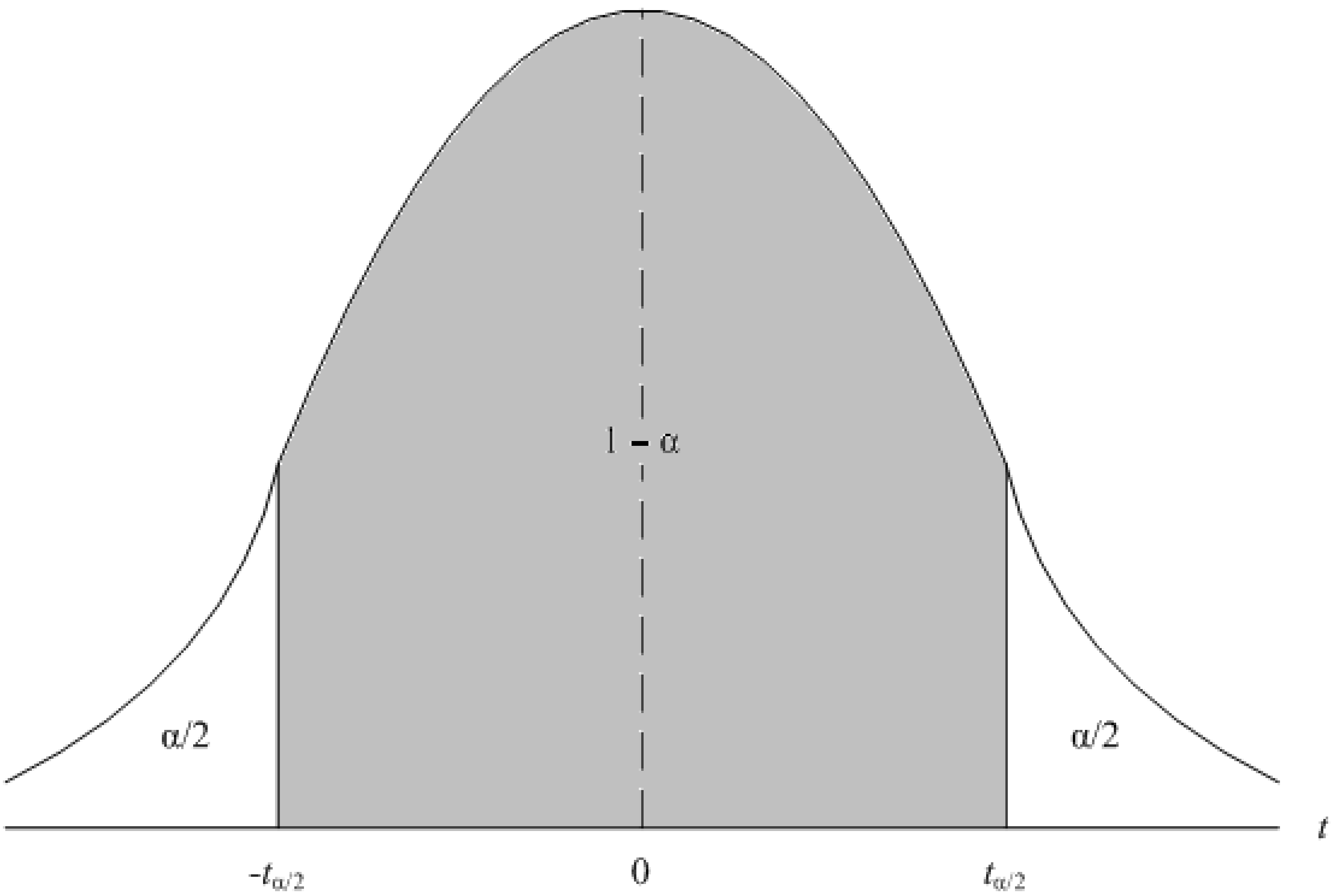
- Con frecuencia no se conoce la varianza de la población, por lo que para estimar la media de la población se recurre a la distribución *t*.
- Al escribir $t_{\alpha/2}$ para el valor *t* el cual se tiene un área de $\alpha/2$ con *v* grados de libertad, se puede ver que

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

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

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

Estimación de la Media
Una Muestra (cont.)



Estimación de la Media Una Muestra (cont.)

- Al multiplicar cada término en la desigualdad por $S/\sqrt{n}$, después restar $\overline{X}$ y multiplicar por -1, se obtiene:

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

- Se selecciona una muestra aleatoria de tamaño *n* de una población cuya varianza no se conoce y se calcula la media y la desviación estándar de la muestra para obtener un intervalo de confianza de (1 – α)100%.

Estimación de la Media
Una Muestra (cont.)

- **Intervalo de Confianza de μ con σ desconocida.** Si $\bar{x}$ y s son la media y la desviación estándar de una muestra aleatoria de tamaño n de una población normal con varianza σ^2, desconocida, un intervalo de confianza de $(1 - \alpha)100\%$ para μ está dado por

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

- Donde $t_{\alpha/2}$ es el valor t con $v = n - 1$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Estimación de la Media Una Muestra (cont.)

- **Teorema.** Si se utiliza $\bar{x}$ como una estimación de μ, se puede tener una confianza de (1 – α)100% de que el error no excederá de $t_{\alpha/2}\, s/\sqrt{n}$.
- **Teorema.** Si se utiliza $\bar{x}$ como una estimación de μ, se puede tener una confianza de (1 – α)100% de que el error no excederá una cantidad específica *e* cuando el tamaño de la muestra es

$$n = \left(t_{\alpha/2}\frac{s}{e}\right)^2 \qquad e = t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Estimación de la Media Una Muestra (cont.)

- Los estadísticos recomiendan que aun cuando no se puede suponer la normalidad, con σ desconocida y $n \geq 30$, s puede remplazar a σ y utilizar el intervalo de confianza

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

- Por lo general, este se denomina como un **intervalo de confianza de muestra grande**.
- La calidad de este enfoque mejora conforme el tamaño de la muestra crece más.

Límites de Confianza Unilaterales

- Los intervalos de confianza y los límites de confianza vistos hasta ahora son en realidad bilaterales (se da tanto el límite inferior como superior).
- Sin embargo, hay muchas aplicaciones en que sólo se requiere un límite.
  - Si la medida de interés es la resistencia a la tensión, el ingeniero recibe más información del límite inferior, escenario del peor caso.
  - Si para la medida de una variable un valor relativamente grande de μ no es provechoso o deseable, entonces resultará de interés el límite superior.

Límites de Confianza Unilaterales (cont.)

- Los límites de confianza unilaterales se desarrollan de la misma forma que los intervalos bilaterales.
- Si $\overline{X}$ es la media de una muestra aleatoria de tamaño n a partir de una población con varianza σ^2, los límites de confianza unilaterales de $(1 - \alpha)100\%$ para μ están dados por

$$\text{límite inferior} \quad \bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

$$\text{límite superior} \quad \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

Límites de Confianza Unilaterales (cont.)

- Si $\overline{X}$ es la media de una muestra aleatoria de tamaño *n* a partir de una población con varianza desconocida, los límites de confianza unilaterales de (1 – α)100% para μ están dados por

$$\text{límite inferior} \quad \bar{x} - t_{\alpha}\frac{s}{\sqrt{n}}$$

$$\text{límite superior} \quad \bar{x} + t_{\alpha}\frac{s}{\sqrt{n}}$$

Error Estándar de una Estimación Puntual

- La desviación estándar de $\overline{X}$ se conoce en la estimación como **error estándar de** $\overline{X}$.
  - De manera simple, el error estándar de un estimador es su desviación estándar.
  - En el caso donde σ se desconoce y el muestreo es sobre una distribución normal, *s* reemplaza a σ y se incluye el **error estándar estimado**.
- El punto importante a considerar es que el ancho del intervalo de confianza de μ depende de la calidad del estimador puntual a través de su error estándar.

Error Estándar de una Estimación Puntual (cont.)

- De esta forma los límites de confianza de μ son

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = \bar{x} \pm z_{\alpha/2}\, s.e.(\bar{x}), \quad \sigma \text{ conocida}$$

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = \bar{x} \pm t_{\alpha/2}\, \hat{s}.\hat{e}.(\bar{x}), \quad \sigma \text{ desconocida}$$

- Donde *s.e.* (*e.e.*) es el error estándar y $\widehat{s.e.}$ ($\widehat{e.e.}$) es el error estándar estimado.

Límites de Predicción

- Algunas veces, aparte de estimar la media de la población, interese predecir los posibles **valores de una observación futura**.
- Este tipo de requerimiento se satisface muy bien mediante la construcción de un **intervalo de predicción**.
- Al predecir una observación futura se necesita la variación de la media y la variación de una observación futura.
  - Por suposición se sabe que la varianza del error aleatorio en una nueva observación es σ^2.

Límites de Predicción (cont.)

- El desarrollo de un intervalo de predicción se representa mejor empezando con una variable aleatoria normal $X_0 - \overline{X}$, donde X_0 es la variable aleatoria de los valores de la nueva observación. Como X_0 y $\overline{X}$ son independientes, entonces

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

$$Z = \frac{X_0 - \overline{X}}{\sqrt{\sigma^2 + \sigma^2/n}} = \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}}$$

$$P\left(-z_{\alpha/2} < \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}} < z_{\alpha/2}\right) = 1-\alpha$$

Límites de Predicción (cont.)

- Para un distribución normal de mediciones con media desconocida μ y varianza conocida σ^2, un **intervalo de predicción** de $(1 - \alpha)100\%$ de una observación futura x_0 es

$$\bar{x} - z_{\alpha/2}\sigma\sqrt{1 + 1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1 + 1/n}$$

- Donde $z_{\alpha/2}$ es el valor z que deja un área de $\alpha/2$ a la derecha.

Límites de Predicción (cont.)

- Resulta importante tratar con el intervalo de predicción de una observación futura en la situación en que se desconoce la varianza de la población; para lo cual se utiliza la distribución *t* de Student en vez de la distribución normal.

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

$$T = \frac{X_0 - \overline{X}}{\sqrt{s^2 + s^2/n}} = \frac{X_0 - \overline{X}}{s\sqrt{1 + 1/n}}$$

$$P\left(-t_{\alpha/2} < \frac{X_0 - \overline{X}}{s\sqrt{1 + 1/n}} < t_{\alpha/2}\right) = 1 - \alpha$$

Límites de Predicción (cont.)

- Para un distribución normal de mediciones con media desconocida μ y varianza desconocida σ^2, un **intervalo de predicción** de $(1 - \alpha)100\%$ de una observación futura x_0 es

$$\bar{x} - t_{\alpha/2} s \sqrt{1 + 1/n} < x_0 < \bar{x} + t_{\alpha/2} s \sqrt{1 + 1/n}$$

- Donde $t_{\alpha/2}$ es el valor t con $v = n - 1$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Límites de Predicción (cont.)

- Los intervalos de predicción unilaterales también son importantes de considerar.
  - Casos donde se debe enfocarse en observaciones futuras grandes se aplican los límites de predicción superiores.
  - Casos donde se debe concentrarse en observaciones futuras pequeñas se sugieren los límites de predicción inferiores.

$$\sigma \text{ conocida} \quad \begin{cases} \text{límite inferior} & \bar{x} - z_\alpha \sigma \sqrt{1 + 1/n} \\ \text{límite superior} & \bar{x} + z_\alpha \sigma \sqrt{1 + 1/n} \end{cases}$$

$$\sigma \text{ desconocida} \quad \begin{cases} \text{límite inferior} & \bar{x} - t_\alpha s \sqrt{1 + 1/n} \\ \text{límite superior} & \bar{x} + t_\alpha s \sqrt{1 + 1/n} \end{cases}$$

Límites de Tolerancia

- En otros casos, el interés se centra en saber donde cae la mayoría de los valores de la población.
- Esto es útil para conocer el desempeño a largo plazo, no en la siguiente observación.
- Un método para establecer el límite deseado consiste en determinar un intervalo de confianza sobre una **proporción fija** de las mediciones.
  - Al visualizar un muestreo aleatorio de una distribución normal con media conocida μ y varianza σ^2; un límite que cubre el 95% de la población de observaciones es $\mu \pm 1.96\sigma$.

Límites de Tolerancia (cont.)

- Lo anterior denota un **intervalo de tolerancia** y, en realidad, es exacta la cobertura de 95% de las observaciones medidas.
- Sin embargo, en la práctica μ y σ rara vez se conocen; por ello, se debe aplicar $\bar{x} \pm ks$, el intervalo es una variable aleatoria y la **cobertura** de una proporción de la población disfrutada por el intervalo no es exacta.
- Como resultado se aplica un intervalo de confianza de (1 – γ) 100% al planteamiento, ya que no se puede esperar que todo el tiempo el intervalo cubra cualquier proporción específica.

Límites de Tolerancia (cont.)

- Para una distribución normal de mediciones con media μ y desviación estándar σ, desconocidas, los **límites de tolerancia bilaterales** están dados por $\bar{x} \pm ks$, donde *k* se determina de modo que se pueda asegurar con una confianza de $(1 - \gamma)$ 100% que los límites dados contienen al menos la proporción $1 - \alpha$ de las mediciones.
- La tabla A.7 da valores de *k* para $1 - \alpha = 0.9, 0.95, 0.99$; $\gamma = 0.05, 0.01$; y para valores seleccionados de *n* de 2 a 1000.

Límites de Tolerancia (cont.)

- Los **límites de tolerancia unilaterales** están dados por

$$\text{límite inferior} \quad \bar{x} - ks$$
$$\text{límite superior} \quad \bar{x} + ks$$

  donde *k* se determina de modo que se pueda asegurar con una confianza de $(1 - \gamma)100\%$ que los límites dados contienen al menos la proporción $1 - \alpha$ de las mediciones.
- La tabla A.7 da valores de *k* para $1 - \alpha = 0.9, 0.95, 0.99$; $\gamma = 0.05, 0.01$; y para valores seleccionados de *n* de 2 a 1000.

Diferencia entre Límites de Confianza, Límites de Predicción y Límites de Tolerancia

- El **intervalo de confianza** sobre la media es útil cuando el analista de datos esté interesado en la **media de la población**.
  - La media de la población se necesita estimar y el intervalo de confianza produce los límites apropiados.
- El **intervalo de tolerancia** está mucho más atento a dónde cae la **mayoría de las observaciones individuales**.
  - ¿Dónde estará la mayor parte de los valores de la población?
- El **intervalo de predicción** se aplica cuando es importante determinar un límite para un **solo valor**.
  - Ni la media ni la ubicación de la mayoría de la población son la cuestión clave, sólo se requiere la ubicación de una sola nueva observación.

Estimación de la Diferencia entre Medias Dos Muestras

- Si se tiene dos poblaciones normales o, a falta de esta, si n_1 y n_2 son suficientemente grandes, se puede establecer un intervalo de confianza para $\mu_1 - \mu_2$ al considerar la distribución muestral de $\overline{X}_1 - \overline{X}_2$.
- La estimación de la diferencia de dos medias de dos poblaciones $\mu_1 - \mu_2$, se hace por medio del teorema del límite central.
- De acuerdo a este teorema se puede establecer las distribuciones muestrales de $\overline{X}_1 - \overline{X}_2$, que está distribuida de forma aproximadamente normal con $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$ y desviación estándar $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$.

Estimación de la Diferencia entre Medias Dos Muestras (cont.)

- Al escribir $z_{\alpha/2}$ para el valor z el cual se tiene un área de $\alpha/2$ a la derecha, se puede ver que

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

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

$$P\left(-z_{\alpha/2} < \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < z_{\alpha/2}\right) = 1-\alpha$$

Estimación de la Diferencia entre Medias Dos Muestras (cont.)

- Al multiplicar cada término en la desigualdad por $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$, después restar $\overline{X}_1 - \overline{X}_2$ y multiplicar por -1, se obtiene:

$$P\left(\left(\overline{X}_1 - \overline{X}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\overline{X}_1 - \overline{X}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = 1 - \alpha$$

- Se selecciona dos muestras aleatorias de tamaño n_1 y n_2 de dos poblaciones cuyas varianzas σ_1^2 y σ_2^2 se conocen y se calcula la diferencia de las medias de las muestras para obtener un intervalo de confianza de $(1 - \alpha)100\%$.

Estimación de la Diferencia entre Medias Dos Muestras (cont.)

- **Intervalo de Confianza de $\mu_1 - \mu_2$ con σ_1 y σ_2 conocidas.** Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones con varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza de $(1 - \alpha)$ 100% para $\mu_1 - \mu_2$ está dado por

$$\left(\bar{x}_1 - \bar{x}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Donde $z_{\alpha/2}$ es el valor z que deja un área de $\alpha/2$ a la derecha.

Estimación de la Diferencia entre Medias Dos Muestras (cont.)

- Para muestras pequeñas que se seleccionan de poblaciones no normales, no se puede esperar que el grado de confianza sea preciso.
- Para muestras de tamaño $n \geq 30$, sin importar la forma de la mayor parte de las poblaciones, la teoría de muestreo garantiza buenos resultados.

Estimación de la Diferencia entre Medias Dos Muestras (cont.)

- **Intervalo de Confianza de $\mu_1 - \mu_2$ con σ_1 y σ_2 iguales pero desconocidas.** Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2, de poblaciones aproximadamente normales con varianzas iguales pero desconocidas, un intervalo de confianza de $(1 - \alpha)100\%$ para $\mu_1 - \mu_2$ está dado por

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- Donde s_p es la estimación de unión de la desviación estándar poblacional y $t_{\alpha/2}$ es el valor t con $v = n_1 + n_2 - 2$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Estimación de la Diferencia entre Medias Dos Muestras (cont.)

- **Intervalo de Confianza de $\mu_1 - \mu_2$ con σ_1 y σ_2 diferentes pero desconocidas.** Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2, de poblaciones aproximadamente normales con varianzas diferentes y desconocidas, un intervalo de confianza de $(1 - \alpha)100\%$ para $\mu_1 - \mu_2$ está dado por

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$v = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left[\left(s_1^2/n_1\right)^2/\left(n_1 - 1\right)\right] + \left[\left(s_2^2/n_2\right)^2/\left(n_2 - 1\right)\right]}$$

- Donde $t_{\alpha/2}$ es el valor t con v grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Estimación de la Diferencia entre Medias Dos Muestras (cont.)

- En el caso de una diferencia entre dos medias, la interpretación se puede extender a una comparación de dos medias:
  - Si el intervalo de confianza de la diferencia $\mu_1 - \mu_2$ es positivo, se infiere que la $\mu_1 > \mu_2$ con poco grado de error.
  - Si el intervalo de confianza de la diferencia $\mu_2 - \mu_1$ es positivo, se infiere que la $\mu_1 < \mu_2$ con poco grado de error.
  - Si el intervalo de confianza de la diferencia $\mu_2 - \mu_1$ puede permitir que $\mu_2 - \mu_1$ sea igual a 0, se infiere que las medias de las poblaciones pueden ser $\mu_1 = \mu_2$ con poco grado de error.

Error Estándar de una Estimación Puntual

- La desviación estándar de $\overline{X}_1 - \overline{X}_2$ se conoce en la estimación como **error estándar de** $\overline{X}_1 - \overline{X}_2$.
  - De manera simple, el error estándar de un estimador es su desviación estándar.
  - En el caso donde σ_1 y σ_2 se desconocen y el muestreo es sobre una distribución normal, s_1 y s_2 reemplaza a σ_1 y σ_2 y se incluye el **error estándar estimado**.
- El punto importante a considerar es que el ancho del intervalo de confianza de μ_1 y μ_2 depende de la calidad del estimador puntual a través de su error estándar.

Error Estándar de una Estimación Puntual (cont.)

- De esta forma los límites de confianza de $\mu_1 - \mu_2$ son

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \left(\bar{x}_1 - \bar{x}_2\right) \pm z_{\alpha/2}\, s.e.\left(\bar{x}_1 - \bar{x}_2\right), \quad \sigma_1 \text{ y } \sigma_2 \text{ son conocidas}$$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2}\, \hat{s}.\hat{e}.\left(\bar{x}_1 - \bar{x}_2\right), \quad \sigma_1 \text{ y } \sigma_2 \text{ son diferentes y desconocidas}$$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2}\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2}\, \hat{s}.\hat{e}.\left(\bar{x}_1 - \bar{x}_2\right), \quad \sigma_1 \text{ y } \sigma_2 \text{ son iguales y desconocidas}$$

- Donde *s.e.* (*e.e.*) es el error estándar y $\widehat{s.e.}$ ($\widehat{e.e.}$) es el error estándar estimado.

Estimación de la Varianza
Una Muestra

- Si se extrae una muestra de tamaño n de una población normal con varianza σ^2, y se calcula la varianza muestral s^2 se obtiene un valor de la estadística S^2.
  - Esta varianza muestral calculada se usará como estimación puntual de σ^2, por ello la estadística S^2 se llama estimador de σ^2.
- Se puede establecer una estimación por intervalos de σ^2 mediante el uso de la estadística

$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

- La estadística X^2 tiene una distribución chi-cuadrada con $v = n - 1$ grados de libertad cuando las muestras se eligen de una población normal.

Estimación de la Varianza Una Muestra (cont.)

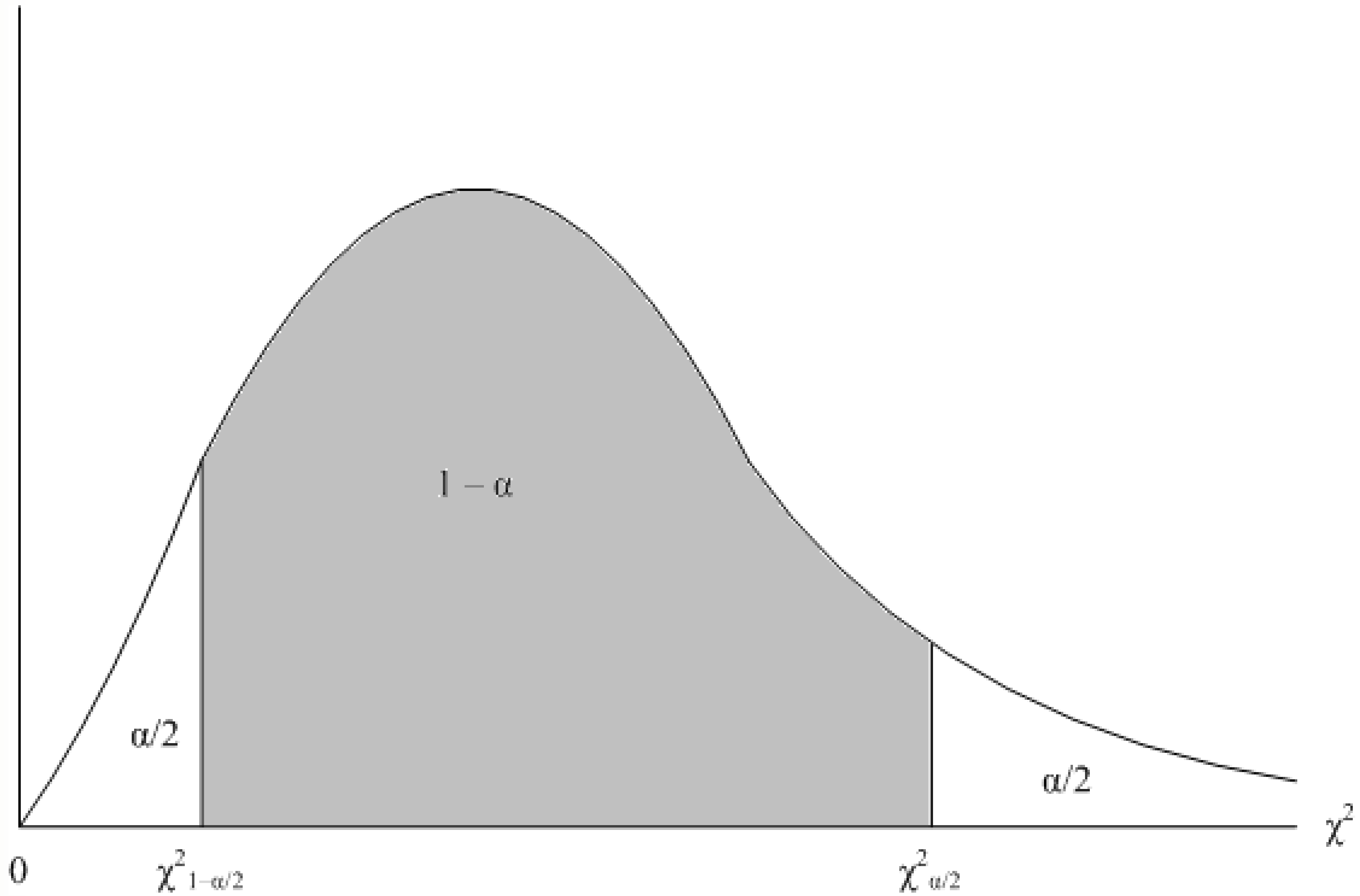
- Se puede escribir:

$$P\left(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}\right) = 1-\alpha$$

$$P\left(\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}\right) = 1-\alpha$$

- Donde $\chi^2_{1-\alpha/2}$ y $\chi^2_{\alpha/2}$ son valores de la distribución chi-cuadrada con $v = n - 1$ grados de libertad, que dejan áreas de $1 - \alpha/2$ y $\alpha/2$, respectivamente, a la derecha.

Estimación de la Varianza
Una Muestra (cont.)



Estimación de la Varianza Una Muestra (cont.)

- Al dividir cada término de la desigualdad entre $(n-1)S^2$, e invertir cada término (lo que cambia el sentido de las desigualdades), se obtiene:

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

- Para la muestra aleatoria particular de tamaño n, se calcula la varianza muestral s^2, y se obtienen el intervalo de confianza de $(1-\alpha)100\%$ para σ^2.

Estimación de la Varianza Una Muestra (cont.)

- **Intervalo de confianza para σ^2.** Si s^2 es la varianza de una muestra aleatoria de tamaño n de una población normal, un intervalo de confianza $(1 - \alpha)100\%$ para σ^2 es:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

- Donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son valores χ^2 con $v = n - 1$ grados de libertad, que dejan áreas de $\alpha/2$ y $1 - \alpha/2$, respectivamente, a la derecha.

Estimación de la Varianza Una Muestra (cont.)

- **Intervalo de confianza para σ.** Un intervalo de confianza $(1 - \alpha)100\%$ para σ se obtiene al tomar la raíz cuadrada de cada extremo del intervalo de confianza para σ^2.

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}$$

- Donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son valores χ^2 con $v = n - 1$ grados de libertad, que dejan áreas de $\alpha/2$ y $1 - \alpha/2$, respectivamente, a la derecha.

Estimación de la Razón de Dos Varianzas Dos Muestras

- Una estimación puntual de la razón de dos varianzas poblacionales σ^2_1/σ^2_2 está dada por la razón s^2_1/s^2_2 de las varianzas muestrales.
  - Esta varianza muestral calculada se usará como estimación puntual de σ^2_1/σ^2_2, por ello la estadística S^2_1/S^2_2 se llama estimador de σ^2_1/σ^2_2.
- Si σ^2_1 y σ^2_2 son las varianzas de poblaciones normales, se puede establecer una estimación por intervalos de σ^2_1/σ^2_2 mediante el uso de la estadística:

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

- La variable aleatoria F tiene una distribución f con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad.

Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)

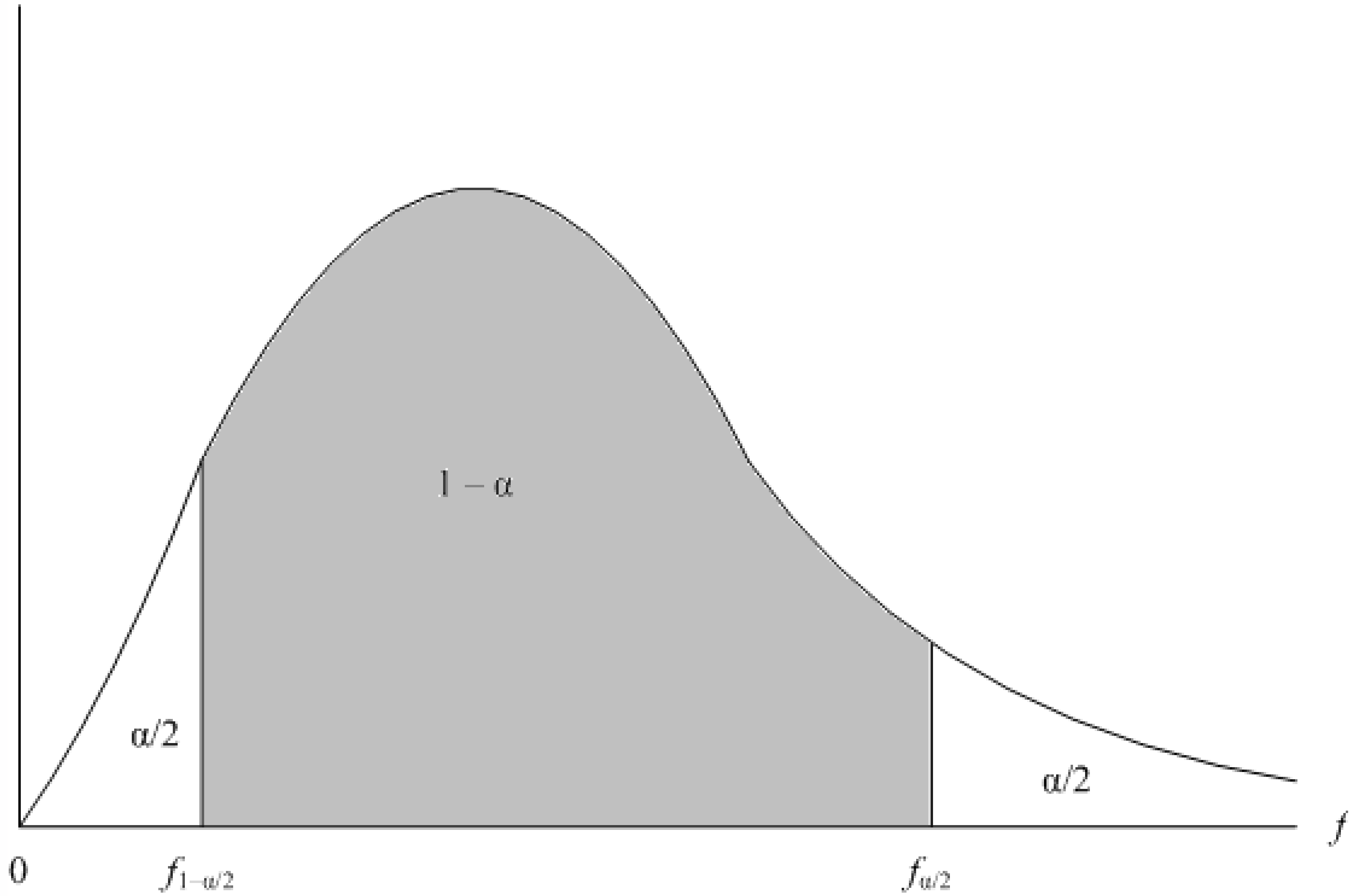
- Se puede escribir:

$$P\left(f_{1-\alpha/2}(v_1,v_2) < F < f_{\alpha/2}(v_1,v_2)\right) = 1-\alpha$$

$$P\left(f_{1-\alpha/2}(v_1,v_2) < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2}(v_1,v_2) \right) = 1-\alpha$$

- Donde $f_{1-\alpha/2}(v_1,v_2)$ y $f_{\alpha/2}(v_1,v_2)$ son valores de la distribución f con v_1 y v_2 grados de libertad, que dejan áreas de $1 - \alpha/2$ y $\alpha/2$, respectivamente, a la derecha.

Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)



Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)

- Al multiplicar cada término en la desigualdad por $S^2{}_2/S^2{}_1$, y después invertir cada término (para cambiar el sentido de las desigualdades), se obtiene:

$$P\left(\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1,v_2)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(v_1,v_2)}\right)=1-\alpha$$

Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)

- Además, se puede reemplazar la cantidad $f_{1-\alpha/2}(v_1,v_2)$ y por $1/f_{\alpha/2}(v_1,v_2)$. Por lo tanto:

$$P\left(\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1,v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2,v_1)\right) = 1-\alpha$$

- Para cualesquiera dos muestras aleatorias independientes de tamaño n_1 y n_2 que se seleccionan de dos poblaciones normales, la razón de las varianzas muestrales $s^2{}_2/s^2{}_1$, se calcula y se obtiene el intervalo de confianza de $(1-\alpha)100\%$ para $\sigma^2{}_1/\sigma^2{}_2$.

Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)

- **Intervalo de confianza para σ^2_1/σ^2_2.** Si s^2_1 y s^2_2 son las varianzas de muestras independientes de tamaño n_1 y n_2, respectivamente, de dos poblaciones normales, entonces un intervalo de confianza $(1 - \alpha)100\%$ para σ^2_1/σ^2_2 es:

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1)$$

- Donde $f_{\alpha/2}(v_1,v_2)$ es un valor de la distribución f con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad, que deja área de $\alpha/2$ a la derecha.

Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)

- **Intervalo de confianza para σ_1/σ_2.** Un intervalo de confianza $(1 - \alpha)100\%$ para σ_1/σ_2 es se obtiene al tomar la raíz cuadrada de cada extremo del intervalo de confianza para $\sigma^2{}_1/\sigma^2{}_2$:

$$\sqrt{\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)}} < \frac{\sigma_1}{\sigma_2} < \sqrt{\frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1)}$$

- Donde $f_{\alpha/2}(v_1,v_2)$ es un valor de la distribución f con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad, que deja área de $\alpha/2$ a la derecha.

Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)

- En el caso de la razón de dos varianzas, la interpretación se puede extender a una comparación de dos varianzas:
  - Si el intervalo de confianza de la razón σ^2_1/σ^2_2 permite la posibilidad de que σ^2_1/σ^2_2 sea igual a 1, se infiere que $\sigma^2_1 = \sigma^2_2$ con poco grado de error.
  - Si el intervalo de confianza de la razón σ^2_1/σ^2_2 no permite la posibilidad de que σ^2_1/σ^2_2 sea igual a 1, se infiere que $\sigma^2_1 \neq \sigma^2_2$ con poco grado de error.

Estimación de la Razón de Dos Varianzas Dos Muestras (cont.)

- En el caso de la razón de dos desviaciones estándar, la interpretación se puede extender a una comparación de dos desviaciones estándar:
  - Si el intervalo de confianza de la razón σ_1/σ_2 permite la posibilidad de que σ_1/σ_2 sea igual a 1, se infiere que $\sigma_1 = \sigma_2$ con poco grado de error.
  - Si el intervalo de confianza de la razón σ_1/σ_2 no permite la posibilidad de que σ_1/σ_2 sea igual a 1, se infiere que $\sigma_1 \neq \sigma_2$ con poco grado de error.

Referencias Bibliográficas

- Walpole, R.E.; Myers, R.H.; Myers, S.L. & Ye, K. "Probabilidad y estadística para ingeniería y ciencias". Octava Edición. Pearson Prentice-Hall. México, 2007.