1 Estimación puntual

Sea la v.a. poblacional X con distribución de probabilidad $f_\theta(x)$, donde θ es el parámetro poblacional desconocido.

Para hacer inferencia sobre el posible valor de $\theta \Rightarrow$

* Seleccionamos una m.a.s. de X, $X_1, X_2, ..., X_n$,

y definimos un **estimador** $T(X_1, X_2, ..., X_n)$ **para** θ utilizando la información muestral.

* El valor concreto que toma el estimador para una realización muestral concreta,

$\widehat{\theta} = T(x_1, x_2, ..., x_n)$, se le llama **estimación puntual de** θ y

es una solución particular a nuestro problema.

Dado que es posible definir muchos estadísticos que sirvan para hacer inferencias sobre el valor del parámetro θ, el objetivo es seleccionar aquellas expresiones que proporcionen garantías en el proceso de estimación paramétrica. Para ello vamos a formular propiedades deseables para la determinación de un buen estiamdor.

1.1 Características deseables de los estimadores

Sea $T(X_1, X_2, ..., X_n)$ un estimador del parámetro θ.

- Diremos que el estimador es **insesgado** si

$$E\left[T(X_1, X_2, ..., X_n)\right] = \theta \Leftrightarrow \text{ El estimador es } \textbf{exacto}$$

Mide la *exactitud* del estimador ya que representa la concentración de las estimaciones entorno al verdadero valor del parámetro.

Cuando $E\left[T(X_1, X_2, ..., X_n)\right] \neq \theta$, se dice que el estimador es **sesgado**.

- Diremos que un estimador insesgado es **consistente** si

$$Var\left[T(X_1, X_2, ..., X_n)\right] \underset{n\to\infty}{\longrightarrow} 0 \Leftrightarrow \text{ El estimador es } \textbf{preciso}$$

Mide la *precición* del estimador ya que representa la concentración de las estimaciones entorno al valor medio del estimador.

Nota: Veamos si los estadísticos media muestral, varianza muestral y proporción muestral verifican las anteriores propiedades deseables:

- Para una variable aleatoria poblacional X con $E(X) = \mu$ y $Var(X) = \sigma^2$, definimos el estadístico **MEDIA MUESTRAL**:

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

que se emplea para hacer inferencias sobre la media poblacional μ.

Se tiene que:

$$E(\overline{X}) = \mu$$
$$Var(\overline{X}) = \frac{\sigma^2}{n}$$
$\Rightarrow$ $\overline{X}$ es un **estimador insesgado** y **consistente** para μ

- Cuando tenemos una realización muestral, la varianza muestral es una medida de la variabilidad de los datos entorno a la media muestral.

Para una variable aleatoria poblacional X con $E(X) = \mu$ y $Var(X) = \sigma^2$, definimos el estadístico **VARIANZA MUESTRAL**

$$\boxed{S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{n}{n-1}\left(\overline{X^2} - \overline{X}^2\right)}$$

que se emplea para hacer inferencias sobre la varianza poblacional σ^2.

Además, hemos visto que para obtener el comportamiento aleatorio de $\overline{X}$ (ya sea aplicando la aditividad del modelo normal o vía el Teorema Central del Límite) es necesario que σ^2 sea conocida. Sin embargo, en la mayoría de los casos prácticos, la varianza σ^2 es desconocida, por lo que es imprescindible estimarla de los datos incluso si únicamente queremos hacer inferencias sobre μ.

Se puede comprobar que:

$$E(S^2) = \sigma^2$$
$$Var(S^2) = \frac{2\sigma^4}{n-1}$$
$\Rightarrow$ S^2 es un **estimador insesgado** y **consistente** para σ^2

- Para una variable aleatoria poblacional X dicotómica, $X \sim b(p)$, definimos el estadístico **PROPORCIÓN MUESTRAL**:

$$\boxed{\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n}}$$

que se emplea para hacer inferencias sobre la proporción de unidades que en la población verifican la propiedad de interés, p.
Se tiene que:

$$E(\widehat{p}) = p$$
$$Var(\widehat{p}) = \frac{p(1-p)}{n}$$

$\Rightarrow$ $\widehat{p}$ es un **estimador insesgado** y **consistente** para p

2 Estimación por intervalos.

En la estimación puntual, calculamos el valor del estimador del parámetro para una muestra concreta e inferimos ese valor a la población. El problema fundamental radica en que, aunque el estimador cumpla las propiedades deseables, no sabemos si la estimación puntual ontenida con él para una muestra concreta estará o no próxima al valor del parémetro ya que éste es desconocido antes y despúes del muestreo. Una desventaja de la estimación puntual es que no podemos establecer el error cometido al estimar el parámetro, ni la fiabilidad de nuestra estimación. En otras palabras, no queremos limitarnos a dar un valor concreto para aproximar un parámetro sino proponer una medida del error que pensamos cometer. Para ello, vamos a proporcionar un intervalo que contenga el valor del parámetro.

`Objetivo:` Basándose en la información contenida en las observaciones muestrales se busca un intervalo de valores dentro del cual se tiene cierta "seguridad" de que se encuentre el valor del parámetro. A este intervalo se le denomina **intervalo de confianza para el parámetro** θ.

`Definición:`

Un intervalo aleatorio (intervalo en el que sus dos extremos son variables aleatorias)

$$I(X_1, X_2, ..., X_n) = [I_1(X_1, X_2, ..., X_n), I_2(X_1, X_2, ..., X_n)]$$

se dice que es un **intervalo de confianza para** θ **al nivel de confianza** $1-\alpha$, con $\alpha \in (0,1)$, si se cumple que

$$P\left(\theta \in [I_1(X_1, X_2, ..., X_n), I_2(X_1, X_2, ..., X_n)]\right) = 1-\alpha$$

2.1 Método para construir intervalos de confianza

1. Encontrar una variable aleatoria $U = h(T, \theta)$, función del parámetro y del estadístico T, cuya distribución de probabilidad sea independiente de θ y de cualquier otro parámetro desconocido.

2. Seleccionar valores $U_1(\alpha)$ y $U_2(\alpha)$ en la distribución de probabilidad de v.a. U tal que

$$P(U_1(\alpha) \leq U \leq U_2(\alpha)) = 1 - \alpha$$

siendo α un número $\in (0, 1)$ y $1 - \alpha$ el nivel de confianza pedido.

3. Despejar adecuadamente el parámetro θ en el intervalo anterior de manera que obtengamos:

$$P(g_1(T, \alpha) \leq \theta \leq g_2(T, \alpha)) = 1 - \alpha$$

2.1.1 Intervalo de confianza nivel $1-\alpha$ para la media μ de una población normal, $X \sim N(\mu, \sigma)$, con σ^2 conocida

Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de una v.a. $X \sim N(\mu, \sigma)$

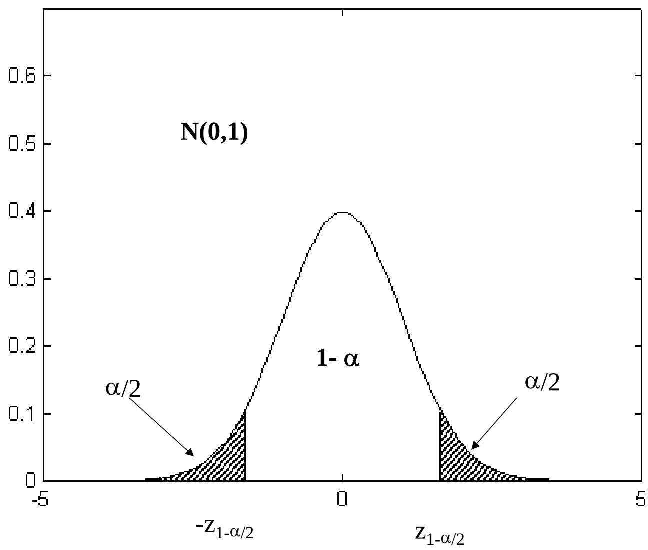
- Para hacer inferencias sobre μ consideramos la media muestral $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, que al tipificarla tenemos que la variable aleatoria

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

- El intervalo será tal que

$$P\left[U_1(\alpha) \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq U_2(\alpha)\right] = 1 - \alpha$$

- Consideramos los valores $U_1(\alpha) = -z_{1-\frac{\alpha}{2}}$ y $U_2(\alpha) = z_{1-\frac{\alpha}{2}}$, ya que en la distribución normal se verifica que:



- Entonces,

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

- Operando adecuadamente obtenemos

$$P(\overline{X} - z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

- Deducimos que un intervalo de confianza al nivel $(1-\alpha)$ para la media poblacional μ es

$$\boxed{[\overline{X} - z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}] = \overline{X} \pm \underbrace{z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}}_{\textbf{MARGEN de ERROR}}}$$

Para cada realización muestral $(x_1, x_2, ..., x_n)$ se obtiene un intervalo de confianza distinto para μ al nivel de confianza $1-\alpha$:

$$\boxed{[\overline{x} - z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}]}$$

<u>Nota:</u>
Cuando la v.a. poblacional X no sea necesariamente normal con $E(X) = \mu$ y $Var(X) = \sigma^2$, el intervalo de confianza obtenido sigue siendo válido para muestras grandes vía el Teorema Central del Límite, ya que de esta manera podemos asegurar que $\overline{X}$ se distribuye aproximadamente según una $N(\mu, \frac{\sigma}{\sqrt{n}})$ cuando n es grande.

2.1.2 Determinación del tamaño muestral cuando σ^2 es conocida

En algún momento en el trabajo estadístico tendremos que decidir qué **tamaño muestral será seleccionado** de la población. Esto es, para estimar la media poblacional μ me planteo el número de observaciones de la muestra son necesarias para garantizar, con una confianza dada que el margen de error sea menor que una cantidad prefijada.
El problema planteado es:

- Dado el nivel de confianza $100(1-\alpha)\%$ y fijado *el margen de error máximo que permito cometer* $\equiv err$.

- Me planteo el valor de n para que:

$$z_{\left(1-\frac{\alpha}{2}\right)} \cdot \frac{\sigma}{\sqrt{n}} \leq err$$

- Despejando obtendremos el valor de n.

$$\boxed{n \geq \left(\frac{z_{\left(1-\frac{\alpha}{2}\right)} \cdot \sigma}{err}\right)^2}$$

Nota.- Debenos tener en cuenta que n debe ser entero.

2.1.3 Intervalo de confianza nivel $1-\alpha$ para la media μ de una población normal, $X \sim N(\mu, \sigma)$, con σ^2 desconocida

Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de una v.a. $X \sim N(\mu, \sigma)$

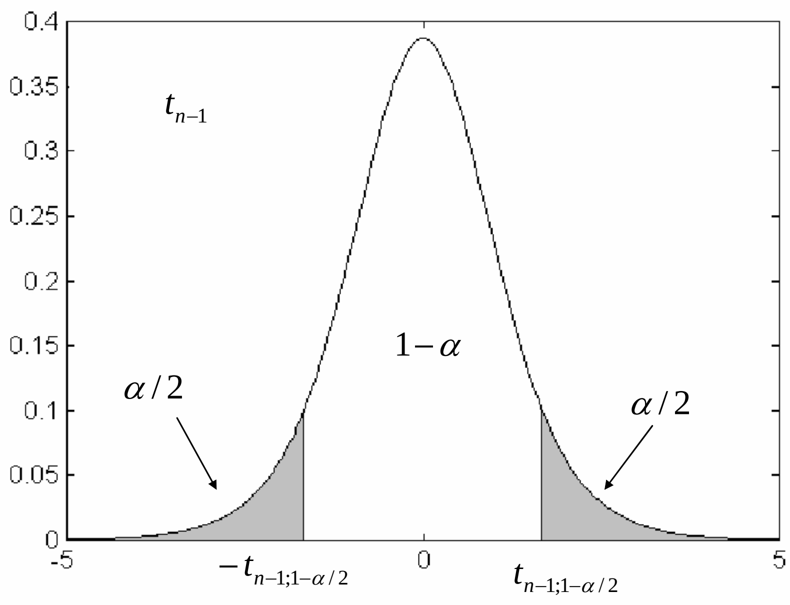
- Consideremos la variable aleatoria

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

- El intervalo será tal que

$$P\left[U_1(\alpha) \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq U_2(\alpha)\right] = 1-\alpha$$

- Consideramos los valores $U_1(\alpha) = -t_{n-1;1-\frac{\alpha}{2}}$ y $U_2(\alpha) = t_{n-1;1-\frac{\alpha}{2}}$, ya que en la distribución t de Student se verifica que:



- Entonces,

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

- Operando adecuadamente obtenemos

$$P(\overline{X} - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}) = 1-\alpha$$

- Deducimos que un intervalo de confianza al nivel $(1-\alpha)$ para la media poblacional μ es:

$$\boxed{[\overline{X} - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}] = \overline{X} \mp \underbrace{t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}}_{\textbf{MARGEN de ERROR}}}$$

Para cada realización muestral $(x_1, x_2, ..., x_n)$ se obtiene un intervalo de confianza distinto para μ al nivel de confianza $1-\alpha$:

$$\boxed{[\overline{x} - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}]}$$

`Interpretación frecuentista de un intervalo de confianza a nivel` $1-\alpha$: Cada vez que tomemos una muestra de tamaño n tendremos que la media poblacional μ se encontrará dentro de dicho intervalo con una confianza del $100(1-\alpha)\%$ y sólo un $100\alpha\%$ de las veces, el intervalo propuesto por este método no recogerá al verdadero valor de μ.

`Nota:` **Factores que influyen en la amplitud de un intervalo de confianza para la media:**

- El nivel de confianza $1-\alpha$: A mayor nivel de confianza mayor amplitud del intervalo
- El tamaño muestral n: A mayor tamaño muestral menor amplitud del intervalo
- Desviación típica de la población σ o estimación de la desviación típica poblacional, S: A mayor σ o S mayor amplitud del intervalo

2.1.4 Intervalo de confianza nivel $1-\alpha$ para la proporción poblacional.

- Sea $(X_1, X_2, ..., X_n)$ una m.a.s. de tamaño n de una v.a. $X_1 \sim B(n,p)$.
- Por tanto cada $X_i \sim b(p)$.(X_i toma el valor 1 con probabilidad p y el valor 0 con probabilidad $1-p$).
- Si denotamos por:

$$\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n} \text{ proporción de } 1's \text{ en la muestra. } \textbf{(proporción muestral)}$$

sabemos que:

$$\frac{\widehat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Por tanto tendríamos que un intervalo de confianza al nivel $(1-\alpha)$ para la proporción poblacional vendrá dado por:

$$\boxed{[\widehat{p}-z_{(1-\frac{\alpha}{2})}\cdot\sqrt{\frac{p(1-p)}{n}}, \widehat{p}+z_{(1-\frac{\alpha}{2})}\cdot\sqrt{\frac{p(1-p)}{n}}]}$$

- Ahora bien p **es desconocido, pero aparece en la parte dedicada a la varianza**. Por tanto, en la práctica se sustituye el término $\sqrt{\frac{p(1-p)}{n}}$ por $\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$ el cuál si se puede obtener a partir de los valores observados.

- Entonces, para cada realización muestral $(x_1, x_2, ..., x_n)$ se obtiene un intervalo de confianza distinto para p al nivel de confianza $1-\alpha$:

$$\boxed{[\widehat{p}-z_{(1-\frac{\alpha}{2})}\cdot\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p}+z_{(1-\frac{\alpha}{2})}\cdot\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}]}$$

2.1.5 Determinación del tamaño muestral para la proporción poblacional.

En este caso debemos determinar qué tamaño de muestra debemos tomar para cumplir unas especificaciones en cuanto al error. Esto es, para estimar la proporción poblacional p, me planteo el número de observaciones de la muestra son necesarias para garantizar, con una confianza dada que el margen de error sea menor que una cantidad prefijada.
El problema planteado es:

- Dado el nivel de confianza $100(1-\alpha)\%$ y fijado *el margen de error máximo que permito cometer* $\equiv err$.

- Me planteo el valor de n para que:

$$z_{(1-\frac{\alpha}{2})}\cdot\sqrt{\frac{p(1-p)}{n}} \leq err$$

- Despejando obtendremos el valor de n.

$$\boxed{n \geq \left(\frac{z_{(1-\frac{\alpha}{2})}\cdot\sqrt{p(1-p)}}{err}\right)^2 = \frac{z^2_{(1-\frac{\alpha}{2})}p(1-p)}{err^2}}$$

Ahora bien, en este caso n depende del parámetro desconocido p.

- Si observamos la función $f(p) = kp(1-p)$ se trata de una parábola que alcanza su máximo para $p = 0.5$. Por tanto una actitud conservadora nos llevaría a tomar ese valor para p y de esta manera garantizamos que con el tamaño muestral obtenido se cumpliría que el error cometido es menor o igual que el prefijado sea cual sea el verdadero valor del parámetro p. En este caso $(p = 0.5)$ tenemos:

$$n \geq \frac{z^2_{(1-\frac{\alpha}{2})}0.5(1-0.5)}{err^2} = \frac{z^2_{(1-\frac{\alpha}{2})}\frac{1}{2}\frac{1}{2}}{err^2}$$

si tomamos $1-\alpha = 0.95 \Rightarrow z_{\left(1-\frac{\alpha}{2}\right)} = 1.96 \simeq 2$ por tanto:

$$\boxed{n \geq \frac{2^2\frac{1}{2}\frac{1}{2}}{err^2} = \frac{1}{err^2}}$$

expresión usual en este tipo de problemas.

- **Nota.- Al igual que en el caso anterior debenos tener en cuenta que n debe ser entero.**

2.2 CASO DE DOS POBLACIONES

Generalicemos lo visto anteriormente al caso de dos poblaciones.

2.2.1 Intervalo de confianza nivel $1-\alpha$ para la diferencia de medias en poblaciones normales con varianzas conocidas.

- Sea $(X_{11}, X_{12}, ..., X_{1n_1})$ una m.a.s. de tamaño n_1 de una v.a.

$$X_1 \sim N(\mu_1, \sigma_1) \Rightarrow \overline{X_1} \sim N(\mu_1, \frac{\sigma_1}{\sqrt{n_1}})$$

- Sea $(X_{21}, X_{22}, ..., X_{2n})$ una m.a.s. de tamaño n_2 de una v.a.

$$X_2 \sim N(\mu_2, \sigma_2) \Rightarrow \overline{X_2} \sim N(\mu_2, \frac{\sigma_2}{\sqrt{n_2}})$$

- De manera análoga al caso de una población se tiene que:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Por tanto un intervalo de confianza al nivel $(1-\alpha)$ para la diferencia entre las medias poblacionales $\mu_1 - \mu_2$ es

$$[(\overline{X}_1 - \overline{X}_2) - z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\overline{X}_1 - \overline{X}_2) + z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}] =$$

$$\boxed{(\overline{X}_1 - \overline{X}_2) \mp \underbrace{z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{\textbf{MARGEN de ERROR}}}$$

2.2.2 Intervalo de confianza nivel $1-\alpha$ para la diferencia de medias en poblaciones normales con varianzas conocidas.

- Sea $(X_{11}, X_{12}, ..., X_{1n_1})$ una m.a.s. de tamaño n_1 de una v.a.

$$X_1 \sim N(\mu_1, \sigma_1) \Rightarrow \overline{X_1} \sim N(\mu_1, \frac{\sigma_1}{\sqrt{n_1}})$$

- Sea $(X_{21}, X_{22}, ..., X_{2n})$ una m.a.s. de tamaño n_2 de una v.a.

$$X_2 \sim N(\mu_2, \sigma_1) \Rightarrow \overline{X_2} \sim N(\mu_2, \frac{\sigma_2}{\sqrt{n_2}})$$

- De manera análoga al caso de una población se tiene que:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Por tanto un intervalo de confianza al nivel $(1-\alpha)$ para la diferencia entre las medias poblacionales $\mu_1 - \mu_2$ es

$$[(\overline{X}_1 - \overline{X}_2) - z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\overline{X}_1 - \overline{X}_2) + z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}] =$$

$$\boxed{(\overline{X}_1 - \overline{X}_2) \mp \underbrace{z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{\textbf{MARGEN de ERROR}}}$$

2.2.3 Intervalo de confianza nivel $1-\alpha$ para la diferencia de medias en poblaciones normales con varianzas desconocidas y distintas.

De manera análoga al caso anterior se tiene que:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_k$$

siendo

$$k = min(n_1 - 1, n_2 - 1).$$

Por tanto un intervalo de confianza al nivel $(1-\alpha)$ para la diferencia entre las medias poblacionales $\mu_1 - \mu_2$ es:

$$[(\overline{X}_1 - \overline{X}_2) - t_{k;(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\overline{X}_1 - \overline{X}_2) + t_{k;(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}] =$$

$$\boxed{(\overline{X}_1 - \overline{X}_2) \mp t_{k;(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

2.2.4 Intervalo de confianza nivel $1-\alpha$ para la diferencia de medias en poblaciones normales con varianzas desconocidas pero asumimos que son iguales.

En este caso se tiene que una buena estimación para la varianza común viene dada por:

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

y en esat situación se tiene:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Por tanto un intervalo de confianza al nivel $(1-\alpha)$ para la diferencia entre las medias poblacionales $\mu_1-\mu_2$ es

$$[(\overline{X}_1 - \overline{X}_2) - t_{n_1+n_2-2;(1-\frac{\alpha}{2})} \cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, (\overline{X}_1 - \overline{X}_2) + t_{n_1+n_2-2;(1-\frac{\alpha}{2})} \cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}]$$

$$\boxed{= (\overline{X}_1 - \overline{X}_2) \mp t_{n_1+n_2-2;(1-\frac{\alpha}{2})} \cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

2.2.5 Intervalo de confianza nivel $1-\alpha$ para la diferencia de dos proporciones.

Sea $(X_{11}, X_{12}, ..., X_{1n_1})$ una m.a.s. de tamaño n_1 de una v.a. $X_1 \sim B(n_1, p_1)$

Sea $(X_{21}, X_{22}, ..., X_{2n})$ una m.a.s. de tamaño n_2 de una v.a. $X_2 \sim B(n_2, p_2)$

De manera análoga al caso de una población se tiene que:

$$\frac{\widehat{p}_1 - \widehat{p}_2 - (p_1-p_2)}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}} \sim N(0,1)$$

Por tanto un intervalo de confianza al nivel $(1-\alpha)$ para la diferencia entre las proporciones poblacionales p_1-p_2 es

$$\boxed{[(\widehat{p}_1 - \widehat{p}_2) - z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}, (\widehat{p}_1 - \widehat{p}_2) + z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}]}$$