TEMA 4.b: Estimación por Intervalos de Confianza

Contents

El objetivo de esta unidad es construir regiones de confianza para una característica desconocida de la población a partir de una muestra aleatoria. Un caso de especial relevancia es el método de obtención de I.C. usando estadísticos pivotales.

4.7 Introducción

En el tema de Estimación Puntual se trata el problema de estimar mediante un número el valor teórico de un parámetro θ desconocido (por ejemplo, la estimación del IPC de un determinado período). Sin embargo, en muchos casos la estimación puntual no es suficiente en el sentido de que no nos indica el error que se comete en la estimación. Lo razonable en la práctica es adjuntar, junto a la estimación puntual del parámetro, un cierto intervalo numérico que mida el margen de error que, de acuerdo a las observaciones muestrales, pueda tener dicha estimación. Surge así la idea de **Intervalo de confianza,** que es un rango de valores entre los que *posiblemente* se encuentre el verdadero valor del parámetro θ.

Definición 1. *La **estimación confidencial o estimación por intervalos** consiste en determinar un posible rango de valores o intervalo, en los que pueda precisarse –con una determinada probabilidad– que el valor de un parámetro se encuentra dentro de esos límites.*

Definición 2. *Dada una muestra aleatoria $X_1, \ldots, X_n$, se denomina **Intervalo de Confianza** para el parámetro θ con **nivel** $1-\alpha$, a un intervalo aleatorio $\left(\widehat{\Theta}_1, \widehat{\Theta}_2\right)$ (cuyos límites dependen de la muestra) de manera que:*

$$P\left(\widehat{\Theta}_1\left(X_1, \ldots, X_n\right) < \theta < \widehat{\Theta}_2\left(X_1, \ldots, X_n\right)\right) = 1-\alpha \;\; \textit{para cada } \theta \in \Theta^{1}$$

*Para una muestra $x_1, \ldots, x_n$, una vez calculados los límites inferior $\widehat{\Theta}_1\left(x_1, \ldots, x_n\right)$ y superior $\widehat{\Theta}_2\left(x_1, \ldots, x_n\right)$, se dice entonces que el intervalo que forman $\left(\widehat{\theta}_1, \widehat{\theta}_2\right)$ constituye una **estimación por intervalo de confianza** de θ con **nivel de confianza** $1-\alpha$ en el sentido de que si se tomasen infinitas muestras $\{x_1, \ldots, x_n\}$ y construyésemos los correspondientes intervalos de confianza, el $100\,(1-\alpha)\,\%$ de ellos contendrían el verdadero valor del parámetro.*

El concepto de "confianza" en el intervalo es muy importante. Supongamos que queremos que el verdadero valor del parámetro θ esté en un intervalo con una confianza del 95%. Esto significa que, si dispusiéramos de todas las muestras posibles, el 95% de ellas contendrían al parámetro, mientras que habrá un 5% de muestras que no lo contendrían. Por tanto, dentro de todas las muestras posibles de una población habrá un 95% de muestras buenas y un 5% de muestras malas. Si en vez de considerar todas las muestras posibles consideramos 100 muestras, por ejemplo, entonces habrá más o menos un 95% de muestras buenas y más o menos un 5% de muestras malas. La figura 1 ilustra la situación. Concretamente han salido 4 intervalos que no contienen al verdadero valor del parámetro y 96 que sí lo contienen. Como por lo general sólo vamos a disponer de una muestra, tenemos que confiar (por ejemplo con un 95% de confianza) que la muestra que tenemos pertenece al grupo de las buenas.

4.8 Obtención de I.C. usando estadísticos pivotales

En general, para construir un intervalo de confianza para un parámetro θ, se utilizará el llamado **método pivotal.**

Definición 3. *Se denomina **estadístico pivotal** al estadístico $T\left(X_1, \ldots, X_n, \theta\right)$ cuya distribución en el muestreo no depende de θ.*

MÉTODO PIVOTAL

Fijado un nivel de confianza $1-\alpha$ $(0 < \alpha < 1)$, el procedimiento general para la construcción de un intervalo de confianza al nivel $1-\alpha$ para un parámetro de interés se desarrolla de acuerdo con el siguiente mecanismo:

1. **Elección del estadístico pivotal.** Se elige un estadístico que dependa solamente del parámetro que se desea estimar y cuya distribución sea conocida (no dependa de θ desconocido).

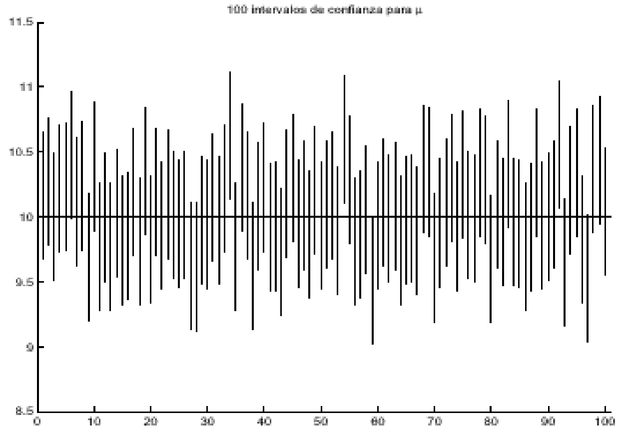

Figure 1:

2. **Planteamiento del enunciado probabilístico.** Se plantea un enunciado probabilístico teniendo en cuenta la distribución de probabilidad del estadístico elegido en la etapa anterior y el valor $1-\alpha$ fijado para el nivel de confianza, es decir, se determinan constantes a y b tales que

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

3. **Transformación del enunciado probabilístico.** Si es posible despejar θ de la expresión anterior, se transforma el enunciado probabilístico en otros equivalentes, mediante operaciones aritméticas, hasta llegar a un enunciado en el que el parámetro de interés figure solo en el centro de la limitación:

$$P\left(T^{-1}\left(X_1, \ldots, X_n; a'\right) < \theta < T^{-1}\left(X_1, \ldots, X_n; b'\right)\right) = 1-\alpha$$ [2]

Entonces, los extremos son los límites de probabidad del enúnciado probabilistico. Dada una realización del vector aleatorio, es decir, $\{x_1, x_2, \cdots, x_n\}$ y sustituyendo estos datos en el intervalo confianza construido anteriormente se obtiene una estimación por intervalo de confianza de θ con un nivel de confianza $1-\alpha$, es decir,

$$\left(T^{-1}\left(x_1, \ldots, x_n; a'\right), T^{-1}\left(x_1, \ldots, x_n; b'\right)\right)$$

A la hora de construir intervalos de confianza surgen una serie de cuestiones:

- **¿Cómo elegir el estadístico pivote?**

  En general, se buscarán funciones sencillas, cuya distribución no dependa del parámetro θ. En los casos más generales, el estadístico pivote surge de la forma más natural; sin embargo, en otras situaciones la búsqueda es más complicada. En estos casos, si tenemos muestras grandes, podemos recurrir como estadístico pivotal al estimador por Máxima Verosimilitud: si $\widehat{\theta}_{MV}$ es el estimador de má-xima verosimilitud de θ, entonces es asintóticamente insesgado, con distribución asintóticamente normal y de varianza asintótica mínima, es decir,

$$\widehat{\theta}_{MV} \text{ es asintóticamente } N\left(\theta, \sigma\left(\widehat{\theta}\right)\right)$$

  de modo que

$$T = \frac{\widehat{\theta}_{MV} - \theta}{\sigma\left(\widehat{\theta}_{MV}\right)} \text{ es asintóticamente } N\left(0,1\right)$$

  es decir, T es el estadístico pivote puesto que depende de θ pero su distribución es conocida (no depende de θ desconocida).

[2]Dependiendo de las operaciones aritméticas a realizar, $a' = a$ y $b' = b$ o $a' = b$ y $b' = a$.

- **¿Cómo determinar las constantes a y b?**

  Interesa coger las constantes a y b de forma que el intervalo tenga longitud mínima. Para ello la distribución del estadístico pivote tiene que ser conocida. Si ésta es unimodal y simétrica, se tomará entonces el I.C. centrado alrededor del valor central, dejando $\alpha/2$ de probabilidad a ambos lados. Si la distribución es asimétrica la selección es bastante complicada, y por sencillez la tomaremos como en el caso anterior, simétricamente.

- **¿Cómo elegir α?**

  Se elegirá α según el nivel de confianza $1-\alpha$ deseado, teniendo en cuenta que, en general, a menor α el intervalo será más largo y por tanto menos informativo. Normalmente se suele tomar como α uno de los siguientes valores: $0'1$, $0'05$ ó $0'01$.

4.9 I.C. para la media y la varianza basado en un muestra para poblaciones normales

$$\begin{array}{c}\textbf{Población normal}\\ X \sim N(\mu,\sigma)\end{array} \left\{ \begin{array}{l} I.C. \text{ para la media } \mu \left\{ \begin{array}{l} \sigma^2 \text{ conocida} \\ \sigma^2 \text{ desconocida} \end{array}\right. \\ I.C. \text{ para la varianza } \sigma^2 \end{array}\right.$$

4.9.1 Intervalo de confianza para la media de una población normal con varianza conocida

Supongamos una población cuya característica en estudio puede describirse mediante una variable aleatoria X con distribución $N(\mu,\sigma^2)$, y sea $X_1,\ldots,X_n$ una muestra aleatoria de la variable X.

Aplicando el procedimiento general para la obtención de un intervalo de confianza a este caso particular:

1. **Elección del estadístico pivote.** En el caso en que conocemos el valor σ de la desviación típica, el estadístico que utilizaremos para la construcción del intervalo de confianza para la media, μ, es

$$\sqrt{n}\frac{\overline{X}-\mu}{\sigma} \sim N(0,1) \text{ con } \overline{X} = \frac{X_1+\ldots+X_n}{n}$$

2. **Planteamiento del enunciado probabilístico.** Si representamos por $z_{\alpha/2}$ el cuantil $\alpha/2$ de la distribución $N(0,1)$, es decir,

$$P\left(Z \leq z_{\alpha/2}\right) = \frac{\alpha}{2}$$

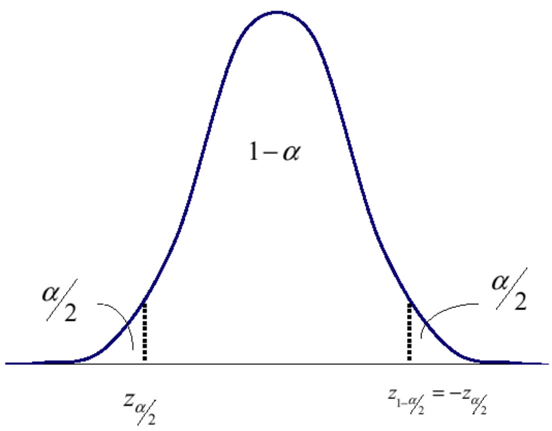

Figure 2:

la probabilidad de que cualquier variable aleatoria con distribución $N(0,1)$ (y en particular el estadístico elegido en la etapa anterior) tome valores en el intervalo $(-z_{1-\alpha/2}, z_{1-\alpha/2})$ es $1-\alpha$ y, por tanto podemos plantear el siguiente enunciado probabilístico:

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

3. **Transformación del enunciado probabilístico**. Si en el enunciado probabilístico de la etapa anterior realizamos sucesivamente las operaciones aritméticas con los términos de las desigualdades del interior del paréntesis

   (a) Multiplicar todos los términos por $\frac{\sigma}{\sqrt{n}}$.

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

   (b) Restar en todos los términos $\overline{X}$.

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

   (c) Multiplicar todos los términos por -1 para cambiar los signos

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

se obtendría finalmente

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

De esta forma hemos construido un intervalo de confianza cuyos extremos son variables aleatorias

$$\boxed{\left(\overline{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)}$$

tales que la probabilidad de que tomen valores entre los que quede comprendido el verdadero valor de la media es $1-\alpha$.

En el supuesto de que dispusiéramos de una realización particular de la muestra $x_1,\ldots,x_n$, evaluando en dicha muestra los extremos del intervalo obtenido, obtendríamos un intervalo de extremos numéricos

$$I = \left(\overline{x}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

que correpondería a una *estimación por intervalo de confianza* del parámetro μ.

Ejemplo 1. *Supongamos que la variable X representa el precio (en miles de pesetas) de la vivienda de alquiler en Madrid y que se distribuye según una normal de media desconocida y varianza conocida* $\sigma^2 = 20^2$. *Para determinar el precio del alquiler medio en Madrid se toma una muestra de 70 viviendas obteniéndose que* $\overline{x} = 82.5$. *Se sabe que la variable*

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

es una normal estándar y por tanto (sabiendo que $z_{1-0.025} = z_{0.975} = 1.96$*)*

$$P\left(-1.96 < \sqrt{n}\frac{\overline{X}-\mu}{\sigma} < 1.96\right) = 0.95$$

Si de esta expresión se despeja el valor μ *se obtiene el intervalo de confianza*

$$0.95 = P\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

Como $\overline{x} = 82.5$ *es la realización particular de* $\overline{X}$*, la estimación por intervalo de confianza de* μ *es*

$$\left(82.5-1.96\frac{20}{\sqrt{70}}, 82.5+1.96\frac{20}{\sqrt{70}}\right) = (77.81, 87.18)$$

Luego el precio del alquiler medio en Madrid se encuentra en el intervalo anterior con una confianza del 95%. Recordar que tenemos que confiar (al 95%) en que este intervalo sea de los "buenos".

Se puede observar que $\overline{x}$ está en el centro del intervalo. La longitud del intervalo es

$$L = 2z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

También se pueden observar los siguientes resultados:

- Cuanto mayor es la desviación típica σ, mayor es la longitud del intervalo.
- Cuanto mayor es el tamaño de la muestra n, menor es la longitud del intervalo.
- Cuanto mayor es el nivel de confianza $1-\alpha$, mayor es la longitud del intervalo.

4.9.2 Intervalo de confianza para la media de una población normal con varianza desconocida

En el caso de que desconozcamos el valor σ de la desviación típica, deberemos estimar dicho parámetro mediante el estimador (insesgado)

$$\hat{S}_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

eligiéndose como estadístico pivotal

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

En general, se procede de forma análoga a la anterior para calcular el intervalo de confianza al $100(1-\alpha)\%$ para la media μ de una población normal con varianza σ^2 desconocida.

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

Si $\overline{x}$ es la realización particular de $\overline{X}$ y $\widehat{s}_X$ es la realización particular de $\hat{S}_X$ para una muestra concreta $\{x_1,\ldots,x_n\}$, la estimación del intervalo de confianza al $100(1-\alpha)\%$ de μ es

$$\boxed{\left(\overline{x} - t_{n-1,1-\alpha/2}\frac{\widehat{s}_X}{\sqrt{n}}, \overline{x} + t_{n-1,1-\alpha/2}\frac{\widehat{s}_X}{\sqrt{n}}\right)}$$

Ejemplo 2. *Supongamos de nuevo que X representa el precio de la vivienda de alquiler en Madrid pero en este caso se desconoce tanto la media como la varianza (caso más habitual que el anterior). Con las mismas 70 viviendas se obtiene* $\overline{x} = 82.5$ *y una cuasivarianza* $\widehat{s}_X^2 = 21.4^2$. *En este caso, se sabe que la variable*

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

es una t de Student con 69 grados de libertad y por tanto (sabiendo que $t_{69,0.975} = 2$*) como* $\overline{x} = 82.5$ *es la realización particular de* $\overline{X}$ *y* $\widehat{s}_X = 21.4$ *es la realización particular de* $\hat{S}_X$*, la estimación por intervalo de confianza de* μ *es*

$$\left(82.5 - 2\frac{21.4}{\sqrt{70}}, 82.5 + 2\frac{21.4}{\sqrt{70}}\right) = (77.38, 87.61)$$

Luego el precio del alquiler medio en Madrid se encuentra en el intervalo anterior con una confianza del 95%.

4.9.3 Intervalo de confianza para la varianza de una población normal

Basándose en la cuasivarianza muestral (estimador insesgado de σ^2)

$$\hat{S}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

se utiliza el siguiente estadístico pivotal

$$\frac{(n-1)\,\hat{S}_X^2}{\sigma^2} \sim \chi_{n-1}^2$$

Se procede como en los anteriores casos, teniendo en cuenta que la distribución de χ_n^2 no es simétrica $\left(\chi^2_{n-1,\alpha/2} \neq -\chi^2_{n-1,1-\alpha/}\right.$

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

Despejando de esta expresión el valor σ^2 se obtiene

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

Si $\hat{s}_X^2$ es la realización particular de $\hat{S}_X^2$ para una muestra concreta $X_1,\ldots,X_n$, la estimación por intervalo de confianza al $100(1-\alpha)\%$ de σ^2 es

$$\boxed{\left(\frac{(n-1)\,\hat{s}_X^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\,\hat{s}_X^2}{\chi^2_{n-1,\alpha/2}}\right)}$$

Ejemplo 3. *Siguiendo con el ejemplo del precio del alquiler en Madrid se tenía que, para 70 viviendas,* $\overline{x} = 82.5$ *y* $\hat{s}_X^2 = 21.4^2$*. Nos preguntamos ahora cuál será la estimación por intervalo de confianza al 95% para la varianza* σ^2 *. En este caso, se sabe que la variable*

$$\frac{(n-1)\,\hat{S}_X^2}{\sigma^2} \sim \chi_{n-1}^2$$

es una χ^2_{69} *con 69 grados de libertad y por tanto, (sabiendo que* $\chi^2_{69,0.025} = 47.92$ *y* $\chi^2_{69,0.975} = 93.86$*)*

$$0.95 = P\left(47.92 < \frac{(n-1)\,\hat{S}_X^2}{\sigma^2} < 93.86\right) = P\left(\frac{(n-1)\,\hat{S}_X^2}{93.86} < \sigma^2 < \frac{(n-1)\,\hat{S}_X^2}{47.92}\right)$$

Como $\hat{s}_X^2 = 21.4^2$ *es la realización particular de* $\hat{S}_X^2$*, la estimación por intervalo de confianza de* σ^2 *es*

$$\left(\frac{(70-1)\,21.4^2}{93.86}, \frac{(70-1)\,21.4^2}{47.92}\right) = (336.86, 659.41)$$

Luego la varianza del precio del alquiler en Madrid se encuentra en el intervalo anterior con una confianza del 95%.

4.10 I.C. para la diferencia de medias y el cociente de varianzas basado en dos muestras para poblaciones normales

Se disponen de dos muestras aleatorias $X_1,\ldots,X_n$ e $Y_1,\ldots,Y_m$ de dos poblaciones normales de interés X e Y, siendo el objetivo construir intervalos que permitan comparar parámetros poblacionales de X e Y.

Poblaciones normales

$$X \sim N(\mu_X, \sigma_X)\,,\; Y \sim N(\mu_Y, \sigma_Y)$$

$$\begin{array}{l} I.C. \text{ para la diferencia } \mu_X - \mu_Y \left\{ \begin{array}{l} \text{Datos indeptes} \left\{ \begin{array}{l} \text{Varianzas conocidas } \sigma_X^2, \sigma_Y^2 \\ \text{Var. desconocidas pero iguales} \\ \text{Var. desconocidas y distintas} \end{array} \right. \\ \text{Datos pareados} \end{array} \right. \\ I.C. \text{ para el cociente de varianzas } \dfrac{\sigma_Y^2}{\sigma_X^2} \end{array}$$

Las muestras que se extraen de las dos poblaciones se pueden obtener por dos procedimientos:

- **Datos pareados.** En este procedimiento los elementos de cada población están relacionados y las muestras se eligen por pares. Se eligen n individuos y se evalúan en ellos conjuntamente X e Y. Por tanto, cada par de variables (X_i, Y_i) corresponde al mismo elemento i de la muestra. Este muestreo reduce notablemente la variabilidad expermental al conseguir grupos más homogéneos.
- **Muestras independientes**. En este procedimiento los elementos de cada población no tienen por qué estar relacionados y las muestras de cada población son independientes entre sí.

4.10.1 Intervalo de confianza para la diferencia de medias con muestras independientes

Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos muestras aleatorias recogidas independientemente de dos poblaciones normales $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu_Y, \sigma_Y)$. Se desea construir intervalos de confianza para la diferencia de medias $\mu_X - \mu_Y$.

- Caso 1: varianzas conocidas

  Puesto que $\overline{X}$ e $\overline{Y}$ son independientes:

$$\left.\begin{array}{l} \overline{X} \sim N\left(\mu_X, \dfrac{\sigma_X}{\sqrt{n}}\right) \\ \overline{Y} \sim N\left(\mu_Y, \dfrac{\sigma_Y}{\sqrt{m}}\right) \end{array}\right\} \Rightarrow \overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$$

  Por tanto

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \sim N(0,1)$$

  puede ser utilizado como estadístico pivote puesto que tiene una distribución conocida independiente de μ_X y μ_Y. Con argumentos idénticos a los realizados en intervalos anteriores se llega a que una estimación por intervalo de confianza de $\mu_X - \mu_Y$ con un nivel de confianza $1 - \alpha$ es

$$\boxed{\left((\overline{x} - \overline{y}) - z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, (\overline{x} - \overline{y}) + z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)}$$

Ejemplo 4. *Supongamos que X representa el precio de la vivienda de alquiler en Madrid e Y representa el precio de la vivienda de alquiler en Barcelona. Supongamos que las varianzas σ_X^2 y σ_Y^2 son conocidas. En este caso,*

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \sim N(0,1)$$

donde n es el tamaño de la primera muestra (los pisos de Madrid) y m es el tamaño de la segunda muestra (los pisos de Barcelona). En este ejemplo de las viviendas $\overline{x} = 82.5$, $\overline{y} = 85.4$, $\sigma_X^2 = 20^2$, $\sigma_Y^2 = 25^2$, $n = 70$ y $m = 50$, de modo que la estimación por intervalo de confianza al 95% de la diferencia de precios medios de alquiler entre Madrid y Barcelona es

$$\left((\overline{x} - \overline{y}) - z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, (\overline{x} - \overline{y}) + z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = (-11.2649, 5.4649)$$

Por tanto, al 95% de confianza no se puede concluir que los precios del alquiler en Barcelona sean más altos que los precios en Madrid, por qué?

- Caso 2: varianzas desconocidas pero iguales

Ahora supongamos que las varianzas σ_X^2 y σ_Y^2 son desconocidas pero iguales (a σ^2). En este caso, se sabe que la variable

$$\overline{X}-\overline{Y}\sim N\left(\mu_X-\mu_Y,\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right)\Leftrightarrow \overline{X}-\overline{Y}\sim N\left(\mu_X-\mu_Y,\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}\right)$$

y por tanto

$$T=\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim N\left(0,1\right)$$

Como en la expresión anterior no se conoce el valor de σ^2, éste se puede sustituir por el valor

$$\widehat{S}^2=\frac{(n-1)\,\widehat{S}_X^2+(m-1)\,\widehat{S}_Y^2}{n+m-2}$$

que es una ponderación de las dos cuasivarianzas muestrales $\widehat{S}_X^2$ y $\widehat{S}_Y^2$. Se puede demostrar que

$$T=\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_Y\right)}{\widehat{S}\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t_{n+m-2}$$

Luego, una estimación por intervalo de confianza de $\mu_X-\mu_Y$ a un nivel de confianza $1-\alpha$ es el siguiente

$$\boxed{\left((\overline{x}-\overline{y})-t_{n+m-2,1-\alpha/2}\widehat{s}\sqrt{\frac{1}{n}+\frac{1}{m}},(\overline{x}-\overline{y})+t_{n+m-2,1-\alpha/2}\widehat{s}\sqrt{\frac{1}{n}+\frac{1}{m}}\right)}$$

Puesto que la construcción de este intervalo requiere suponer que las varianzas, aunque desconocidas, son iguales, es necesario realizar previamente un contraste de hipótesis para garantizar dicha suposición.

Ejemplo 5. *Supongamos de nuevo que X representa el precio de la vivienda de alquiler en Madrid e Y representa el precio de la vivienda de alquiler en Barcelona. Supongamos que las varianzas σ_X^2 y σ_Y^2 son desconocidas pero iguales. Se obtenían los siguientes datos: $\overline{x}=82.5$, $\overline{y}=85.4$, $\widehat{s}_X^2=21.4^2$, $\widehat{s}_Y^2=24.3^2$, $n=70$ y $m=50$. La cuasivarianza ponderada es*

$$\widehat{s}^2=\frac{(n-1)\,\widehat{s}_X^2+(m-1)\,\widehat{s}_Y^2}{n+m-2}=\frac{69\cdot\widehat{s}_X^2+49\cdot\widehat{s}_Y^2}{118}=512.9936$$

Por tanto, la estimación por intervalo de confianza de la diferencia de precios en el alquiler entre Madrid y Barcelona al 95% de confianza es

$$\left((\overline{x}-\overline{y})\pm t_{n+m-2,1-\alpha/2}\widehat{s}\sqrt{\frac{1}{n}+\frac{1}{m}}\right)=(-11.2038,5.4038)$$

Tampoco se puede concluir con un 95% de confianza que los precios del alquiler en Barcelona sean más altos que los de Madrid.

- Caso 3: varianzas desconocidas y distintas

Si las varianzas no se pueden suponer iguales, el estadístico anterior no se puede utilizar. En el caso de que las varianzas fuesen conocidas, el estadístico a utilizar era

$$T=\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\sim N\left(0,1\right)$$

a) **Tamaños muestrales grandes ($n,m\geq 30$)**

En este caso se utiliza el estadístico pivote anterior estimando σ_X por $\widehat{S}_X$ y σ_Y por $\widehat{S}_Y$, es decir,

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_Y\right)}{\sqrt{\frac{\widehat{S}_X^2}{n} + \frac{\widehat{S}_Y^2}{m}}} \underset{A \sin t}{\sim} N(0,1)$$

que tiene una distribución aproximadamente normal $N(0,1)$.

b) **Tamaños muestrales pequeños ($n, m < 30$)**
En este caso la aproximación normal no es nada precisa. La solución más usada habitualmente es la aproximación debida a Welch, según la cual el estadístico anterior sigue una distribución t de Student con $g = n + m - 2 - \delta$ grados de libertad, con δ el entero más próximo a

$$\Delta = \frac{\left[(m-1)\frac{\widehat{S}_X^2}{n} - (n-1)\frac{\widehat{S}_Y^2}{m}\right]^2}{(m-1)\left(\frac{\widehat{S}_X^2}{n}\right)^2 + (n-1)\left(\frac{\widehat{S}_Y^2}{m}\right)^2}$$

Se comprueba que $0 \leq \delta \leq \max(n-1, m-1)$

4.10.2 Intervalo de confianza para la diferencia de medias con datos pareados

La principal característica del muestreo apareado es que ambas muestras son claramente dependientes, de modo que los estadísticos pivote usados en caso de independencia no se pueden emplear en este caso, puesto que los intervalos de confianza pueden salir excesivamente grandes o pequeños. El motivo es que si las variables X e Y son dependientes

$$Var\left(\overline{X} - \overline{Y}\right) = Var\left(\overline{X}\right) + Var\left(\overline{Y}\right) - 2Cov\left(\overline{X}, \overline{Y}\right)$$

con lo que el denominador de los estadísticos pivote puede ser equivocadamente grande o pequeño, según sea la dependencia entre X e Y, es decir, según sea $Cov\left(\overline{X}, \overline{Y}\right)$. La solución es considerar una nueva variable aleatoria $D = X - Y$,

$$\left.\begin{array}{l} X \sim N(\mu_X, \sigma_X) \\ Y \sim N(\mu_Y, \sigma_Y) \end{array}\right\} \Rightarrow Z = X - Y \sim N\left(\mu_Z = \mu_X - \mu_Y, \sqrt{\sigma_Z^2} = \sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}\right)$$

es decir

$$\frac{Z - \mu_Z}{\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}} \sim N(0,1)$$

Puesto que σ_Z es desconocido, si estimamos su valor por la cuasivarianza muestral $\widehat{S}_Z$, el estadístico pivote en este caso es

$$\sqrt{n}\frac{Z - \mu_Z}{\widehat{S}_Z} \sim t_{n-1}$$

El intervalo de confianza para $\mu_X - \mu_Y$ equivale al intervalo para μ_Z y para ello disponemos de una muestra aleatoria de la variable Z sin más que considerar los valores $Z_i = X_i - Y_i$, obtenidos a partir de las muestras apareadas $\{(X_i, Y_i)\}_{i=1}^n$. Recordando la construcción de intervalos de confianza para la media de una población normal, se obtiene que el intervalo de confianza con nivel $1 - \alpha$ para $\mu_X - \mu_Y = \mu_Z$ es

$$\boxed{\left(\left(\overline{X} - \overline{Y}\right) - t_{n-1,1-\alpha/2}\frac{\widehat{S}_Z}{\sqrt{n}}, \left(\overline{X} - \overline{Y}\right) + t_{n-1,1-\alpha/2}\frac{\widehat{S}_Z}{\sqrt{n}}\right)}$$

Ejemplo 6. *Supongamos que X representa el salario de un ejecutivo antes de hacer un MBA e Y representa el salario de un ejecutivo después de hacer un MBA. La siguiente tabla muestra los salarios correspondientes a 7 ejecutivos (en millones de pesetas anuales):*

x_i	2.7838	2.1672	3.0627	3.1428	2.4268	3.5955	3.5946
y_i	3.8962	3.8169	4.1051	3.9720	4.1033	3.9686	5.2383
d_i	-1.1125	-1.6497	-1.0424	-0.8281	-1.6765	-0.3731	-1.6437

La última fila de la tabla representa la variable $Z = X - Y$ .En el caso de los 7 ejecutivos, se obtiene que $\bar{z} = -1.1894$ y $\hat{s}_Z^2 = 0.2466$. Sabiendo que $t_{6,0.975} = -t_{6,0.025} = 2.447$, la estimación por intervalo de confianza de $\mu_Z = \mu_X - \mu_Y$ es

$$\left((\bar{x}-\bar{y}) - t_{n-1,1-\alpha/2}\frac{\hat{s}_Z}{\sqrt{n}}, (\bar{x}-\bar{y}) + t_{n-1,1-\alpha/2}\frac{\hat{s}_Z}{\sqrt{n}}\right) =$$
$$\left(-1.1894 - 2.447\frac{\sqrt{0.2466}}{\sqrt{7}}, -1.1894 + 2.447\frac{\sqrt{0.2466}}{\sqrt{7}}\right) = (-1.648, -0.730)$$

Luego la diferencia de los salarios medios de los ejecutivos antes y después de hacer un MBA se encuentra en el intervalo anterior con una confianza del 95%. Se observa que esta diferencia es menor que 0 luego se puede afirmar, con un 95% de confianza, que el salario medio de un ejecutivo antes de hacer el master es menor que después de hacerlo.

4.10.3 Intervalo de confianza para el cociente de varianzas

Teniendo en cuenta la normalidad e independencia de ambas poblaciones, se verifica que

$$\left.\begin{array}{l} \dfrac{(n-1)\,\hat{S}_X^2}{\sigma_X^2} \sim \chi_{n-1}^2 \\ \dfrac{(m-1)\,\hat{S}_Y^2}{\sigma_Y^2} \sim \chi_{m-1}^2 \end{array}\right\} \Rightarrow \frac{\frac{\hat{S}_X^2}{\sigma_X^2}}{\frac{\hat{S}_Y^2}{\sigma_Y^2}} = \frac{\hat{S}_X^2}{\sigma_X^2}\frac{\sigma_Y^2}{\hat{S}_Y^2} \sim F_{n-1,m-1}$$

Por tanto, un intervalo de confianza a nivel $1-\alpha$ se obtiene de la siguiente manera

$$1-\alpha = P\left(F_{n-1,m-1,\alpha/2} < \frac{\hat{S}_X^2}{\sigma_X^2}\frac{\sigma_Y^2}{\hat{S}_Y^2} < F_{n-1,m-1,1-\alpha/2}\right)$$
$$1-\alpha = P\left(F_{n-1,m-1,\alpha/2}\frac{\hat{S}_Y^2}{\hat{S}_X^2} < \frac{\sigma_Y^2}{\sigma_X^2} < F_{n-1,m-1,1-\alpha/2}\frac{\hat{S}_Y^2}{\hat{S}_X^2}\right)$$
$$1-\alpha = P\left(\frac{\frac{\hat{S}_X^2}{\hat{S}_Y^2}}{F_{n-1,m-1,\alpha/2}} > \frac{\sigma_X^2}{\sigma_Y^2} > \frac{\frac{\hat{S}_X^2}{\hat{S}_Y^2}}{F_{n-1,m-1,1-\alpha/2}}\right)$$

El intervalo de confianza es por lo tanto

$$\boxed{\left(\frac{\frac{\hat{s}_X^2}{\hat{s}_Y^2}}{F_{n-1,m-1,1-\alpha/2}}, \frac{\frac{\hat{s}_X^2}{\hat{s}_Y^2}}{F_{n-1,m-1,\alpha/2}}\right)}$$

Con frecuencia se toma como extremo inferior del intervalo el 0. Así, el intervalo de confianza a nivel $1-\alpha$ es

$$\left(0, \frac{\frac{\hat{s}_X^2}{\hat{s}_Y^2}}{F_{n-1,m-1,\alpha/2}}\right)$$

Ejemplo 7. *Siguiendo con el ejemplo de los alquileres en Madrid y en Barcelona, vamos a obtener un intervalo de confianza para el cociente de varianzas. Se obtenían los siguientes datos:* $\overline{x} = 82.5$, $\overline{y} = 85.4$, $\widehat{s}_X^2 = 21.4^2$, $\widehat{s}_Y^2 = 24.3^2$, $n = 70$ *y* $m = 50$. *Si* $F_{69,49,0.025} = 0.586$ *y* $F_{69,49,0.975} = 1.7348$, *la estimación por intervalo de confianza de* $\frac{\sigma_Y^2}{\sigma_Y^2}$ *es el siguiente:*

$$\left(\frac{\frac{\widehat{s}_X^2}{\widehat{s}_Y^2}}{F_{n-1,m-1,1-\alpha/2}}, \frac{\frac{\widehat{s}_X^2}{\widehat{s}_Y^2}}{F_{n-1,m-1,\alpha/2}} \right) = \left(\frac{\frac{21.4^2}{24.3^2}}{1.7348}, \frac{\frac{21.4^2}{24.3^2}}{0.586} \right) = \left(\frac{0.7756}{1.7348}, \frac{0.7756}{0.586} \right) = (0.4470, 1.3235)$$

4.11 I.C. para muestras no normales y muestras grandes

En esta sección se estudiarán otro tipo de distribuciones y se supondrá que el tamaño de la muestra es suficientemente grande.

4.11.1 Intervalo de confianza para una proporción

Se desea construir un intervalo de confianza para la proporción p de elementos de una población que tienen una determinada característica de interés. Se toma una muestra aleatoria simple $X_1, \ldots, X_n$ de elementos, anotando 1 si dicho elemento tiene la característica de interés y 0 si carece de ella

$$X_i = \begin{cases} 1 & \text{(si tiene la característica) con probabilidad } p \\ 0 & \text{(si no la tiene) con probabilidad } 1-p \end{cases}$$

Disponemos por tanto de una m.a.s. de una variable binomial $Bi(1, p)$. El estimador de máxima verosimilitud de la proporción poblacional p es la proporción muestral

$$\widehat{p} = \frac{X_1 + \cdots + X_n}{n}$$

Utilizando el Teorema Central del Límite, se tiene que

$$T = \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \text{ es una } N(0,1) \text{ aproximadamente}$$

y por tanto, el intervalo de confianza al $100(1-\alpha)\%$ para p es

$$\boxed{\left(\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}, \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \right)}$$

Este intervalo no es calculable al desconocerse p. En la práctica, se sustituye p por su estimador $\widehat{p}$, con lo que obtendríamos como intervalo

$$\left(\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \right)$$

Ejemplo 8. *Estamos interesados en conocer la proporción p de jóvenes que fuman. Sea X la variable que vale 1 si una persona fuma y vale 0 si no.Si se preguntó a 344 estudiantes si fumaban o no y 83 dijeron que sí, el intervalo de confianza al 90% para la verdadera proporción de estudiantes que fuman es, sabiendo que* $z_{1-0.1/2} = z_{0.95} = 1.645$, *y que* $\widehat{p} = \frac{83}{344} = 0.24128$

$$\left(\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \right) =$$
$$\left(0.24128 - 1.645\sqrt{\frac{0.24128(1-0.24128)}{344}}, 0.24128 + 1.645\sqrt{\frac{0.24128(1-0.24128)}{344}} \right) =$$
$$(0.203, 0.279)$$

4.11.2 Intervalo de confianza para la diferencia de dos proporciones

La manera de proceder es idéntica a la del caso de una sola proporción. Se recogen sendas muestras aleatorias simples $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ de las variables X e Y donde

$$X_i = \begin{cases} 1 & \text{con probabilidad } p_X \\ 0 & \text{con probabilidad } 1 - p_X \end{cases} \quad \text{e } Y_i = \begin{cases} 1 & \text{con probabilidad } p_Y \\ 0 & \text{con probabilidad } 1 - p_Y \end{cases}$$

Supuesta aceptable la aproximación normal, se tiene que

$$\left.\begin{array}{l} \dfrac{\widehat{p}_X - p_X}{\sqrt{\dfrac{p_X(1-p_X)}{n}}} \sim N(0,1) \\ \dfrac{\widehat{p}_Y - p_Y}{\sqrt{\dfrac{p_Y(1-p_Y)}{m}}} \sim N(0,1) \end{array}\right\} \Rightarrow \widehat{p}_X - \widehat{p}_Y \sim N\left(p_X - p_Y, \sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}\right)$$

Mediante razonamientos análogos a los seguidos a lo largo del tema, se llega a que el intervalo de confianza para la diferencia de proporciones $p_X - p_Y$ con un nivel de confianza aproximado de $1 - \alpha$ es

$$\boxed{\left((\widehat{p}_X - \widehat{p}_Y) \pm z_{1-\alpha/2}\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}\right)}$$

De nuevo, se ha de sustituir p_X y p_Y por sus estimaciones $\widehat{p}_X$ y $\widehat{p}_Y$ respectivamente.

Ejemplo 9. *Estamos ahora interesados en conocer si existen diferencias entre la proporción p_1 de hombres que fuman y la proporción p_2 de mujeres que fuman. Sea X la variable que vale 1 si un hombre fuma y vale 0 si no, e Y la variable que vale 1 si una mujer fuma y vale 0 si no. Si dentro de los 344 estudiantes había 160 hombres de los cuales 29 fumaban, y 184 mujeres de las cuales 54 fumaban, el intervalo de confianza para la diferencia de proporciones al 90% de confianza es el siguiente*

$$\left((\widehat{p}_X - \widehat{p}_Y) \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_X(1-\widehat{p}_X)}{n} + \frac{\widehat{p}_Y(1-\widehat{p}_Y)}{m}}\right) =$$

$$\left(\left(\frac{29}{160} - \frac{54}{184}\right) \pm 1.645\sqrt{\frac{\frac{29}{160}\left(1 - \frac{29}{160}\right)}{160} + \frac{\frac{54}{184}\left(1 - \frac{54}{184}\right)}{184}}\right) =$$

$$(-0.1868, -0.0377)$$

que evidencia (al 90% de confianza) que la proporción de estudiantes fumadoras es mayor que la proporción de estudiantes fumadores.

4.12 Determinación del tamaño muestral

En algunas situaciones es interesante fijar por adelantado el tamaño del intervalo de confianza. En dichas situaciones se puede calcular el tamaño muestral necesario para garantizar dicha longitud.

Ejemplo 10. *Hemos visto que el intervalo de confianza al $100(1-\alpha)\%$ para la media μ de una población normal (con varianza conocida) es*

$$\left(\bar{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Se puede observar que

$$L = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

es la mitad de la longitud del intervalo, llamada amplitud. Si fijamos esta longitud de antemano, entonces podemos despejar n de la ecuación anterior:

$$n = \frac{z^2_{1-\alpha/2}\sigma^2}{L^2}$$

y este tamaño muestral garantiza que la mitad del intervalo sea L. Por ejemplo, ¿cuántos alquileres de viviendas tenemos que observar para obtener un intervalo de confianza al 95% con una amplitud L menor que 6? La respuesta es

$$n = \frac{1.96^2 20^2}{6^2} = 43.5424$$

que se redondea por exceso a n = 44.

Ejemplo 11. *Se quiere realizar una encuesta de opinión para las futuras elecciones generales. La encuesta cuenta con un margen de error del 3% que quiere decir que la amplitud del intervalo al 95% tiene que ser L = 0.03 . ¿A cuantas personas se tiene que entrevistar para llevar a cabo una encuesta de estas características?*

Sabemos que

$$L = z_{1-\alpha/2}\sqrt{\frac{\widehat{p}\,(1-\widehat{p})}{n}}$$

y de esta fórmula no podemos despejar n porque depende de la proporción muestral que no se conoce de antemano. En el peor de los casos la proporción muestral será $\widehat{p} = 0.5$ *y por tanto, en el peor de los casos, la amplitud del intervalo será*

$$L = z_{1-\alpha/2}\sqrt{\frac{0.5\,(1-0.5)}{n}} = \frac{z_{1-\alpha/2}0.5}{\sqrt{n}}$$

y ahora sí se puede despejar el tamañao muestral:

$$n = \frac{0.25 \cdot z^2_{1-\alpha/2}}{L^2}$$

En el caso de las encuestas hacen falta

$$n = \frac{0.25 \cdot 1.96^2}{0.03^2} \simeq 1068$$

personas para conseguir un margen de error en la encuesta del 3%.

Part I

Tabla de estadísticos pivotales

Estadísticos pivotales para una muestra aleatoria $\{X_1, X_2, \cdots, X_n\}$ de una población Normal:

Para	Suposición	Estadístico Pivotal	Distribución
μ	σ^2 conocida	$\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$	$N(0,1)$
μ	σ^2 desconocida	$\sqrt{n}\frac{\overline{X}-\mu}{\hat{S}_X}$	t_{n-1}
σ^2	μ conocida	$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$	χ^2_n
σ^2	μ desconocida	$\frac{(n-1)\hat{S}^2_X}{\sigma^2}$	χ^2_{n-1}

$$\hat{S}^2_X = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Estadísticos pivotales para muestras grandes (distribución aproximada):

- Estadístico pivotal para la media:

μ	σ^2 desconocida	$\sqrt{n}\frac{\overline{X}-\mu}{\hat{S}_X}$	$N(0,1)$

- Estadístico pivotal para una proporción:

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

- De forma general, dado un estadístico $\hat{\theta}$ para θ con distribución aproximadamente Normal (por ejemplo el estimador Máximo verosímil):

θ	$\frac{\widehat{\theta}-\theta}{\sqrt{var(\widehat{\theta})}}$	$N(0,1)$

Estadísticos pivotales para dos muestras aleatorias independientes $\{X_1, X_2, \cdots, X_n\}$ y $\{Y_1, Y_2, \cdots, Y_m\}$ de una población Normal:

Para	Suposición	Estadístico Pivotal	Distribución
$\mu_X - \mu_Y$	σ_X^2 y σ_Y^2 conocidas	$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}$	$N(0,1)$
$\mu_X - \mu_Y$	$\sigma_X^2 = \sigma_Y^2$ desconocidas	$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\widehat{S}\sqrt{\frac{1}{n}+\frac{1}{m}}}$ $\widehat{S}^2 = \frac{\sum_{i=1}^n (X_i-\bar{X})^2+\sum_{j=1}^m (Y_j-\bar{Y})^2}{n+m-2}$	t_{n+m-2}
$\mu_X - \mu_Y$	$\sigma_X^2 \neq \sigma_Y^2$ desconocidas	$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\widehat{S}_X^2}{n}+\frac{\widehat{S}_Y^2}{m}}}$	$t_{n+m-2-g}$*
$\frac{\sigma_X^2}{\sigma_Y^2}$		$\frac{\widehat{S}_X^2}{\widehat{S}_Y^2}\frac{\sigma_Y^2}{\sigma_X^2}$	$F_{n-1,m-1}$

con $\widehat{S}_X^2$ $\widehat{S}_Y^2$ las cuasivarianzas de cada muestra.

*donde g es el entero más próximo a $\Delta = \frac{\left[(m-1)\frac{\widehat{S}_X^2}{n}-(n-1)\frac{\widehat{S}_Y^2}{m}\right]^2}{(m-1)\left(\frac{\widehat{S}_X^2}{n}\right)^2+(n-1)\left(\frac{\widehat{S}_Y^2}{m}\right)^2}$

Estadístico pivotal para dos muestras aleatorias pareadas $\{(X_i, Y_i)_{i=1}^n\}$ de una población Normal:

Para		Estadístico Pivotal	Distribución
$\mu_X - \mu_Y$	$Z_i = X_i - Y_i$	$\sqrt{n}\frac{\bar{Z}-\mu_Z}{\widehat{S}_Z}$	t_{n-1}

Estadísticos pivotales para muestras grandes (distribuciones aproximadas):

•Estadístico pivotal para la diferencia de medias entre dos poblaciones independientes:

$\mu_X - \mu_Y$	σ_X^2 y σ_Y^2 desconocidas	$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\widehat{S}_X^2}{n}+\frac{\widehat{S}_Y^2}{m}}}$	$N(0,1)$

•Estadístico pivotal para la diferencia entre dos proporciones:

$p_X - p_Y$	$\frac{(\widehat{p}_X-\widehat{p}_Y)-(p_X-p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n}+\frac{p_Y(1-p_Y)}{m}}}$	$N(0,1)$