

A. Parámetros de estadísticas de poblaciones y muestras

Muestreo Aleatorio:

Todos los individuos de una población pueden formar parte de una muestra. En un estudio, las observaciones pueden ser de cualquier tamaño y los resultados pueden ser de forma numérica o descriptiva. Evita el sesgo de datos.

Población:

Total de observaciones en las que el estudio se interesa. Sus elementos pueden tener el nombre de individuos o unidades estadísticas, y el total de estos se llama tamaño de la población.

Variable Estadística:

Hay dos tipos cualitativa y cuantitativa, la cualitativa no se expresa por medio de números sino por descripciones, la cuantitativa por otro lado se representa por medio de números y puede ser discretos o continuos.

Muestra:

Subconjunto de una población. Puede elegirse de forma aleatoria lo cual evita el sesgo de datos o tener un foco específico.

B. Estadística descriptiva (media, mediana, varianza, rango e intercuartiles de una muestra) y gráfica (histogramas, diagramas de caja y diagramas de dispersión).

Estadística Descriptiva:

Media: $\bar{X} = \sum_{i=1}^n \frac{X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$ El promedio de los datos recopilados.

Mediana: $\tilde{X} = \begin{cases} X_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{X_{n/2} + X_{(n/2)+1}}{2} & \text{si } n \text{ es par} \end{cases}$ Parte el conjunto de datos en 2 exactamente en el centro.

Varianza: Se define como el promedio de los cuadrados de la desviación de las observaciones de la media.

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

Rango: Sirve para medir la variabilidad entre la observación superior de la Inferior, principalmente se utiliza para definir intervalos.

$$X_{\max} - X_{\min}$$

Intercuartiles: Los cuartiles son los valores que dividen los datos en 4 partes iguales, los intercuartiles se refieren a los datos que se encuentran entre dos cuartiles diferentes.

Estadística Grafica:

Histogramas: Similares a un diagrama de barras en el que las barras pueden estar pegadas unas con otras y el área de cada rectángulo es respectivo a la frecuencia relativa.

Diagramas de caja: Los datos se representan mediante una caja, dentro de ella se dibuja una línea la cual representa la mediana de los datos, de ella se extienden líneas que van hasta los datos minimos y maximos y dentro de estas líneas se representan datos fuera del rango intercuartil.

Diagramas de dispersión: Se representan los datos de dos variables y se hacen predicciones a raíz de ella, en él cada dato es representado por un punto ubicado en un sistema de coordenadas.

C. Estimación de Parámetros, Intervalos de confianza, estimación de desviaciones estándar

La estadística inferencial, que se refiere al proceso donde se realizan inferencias y generalizaciones de una población, se divide en dos áreas: La estimación y la prueba de hipótesis.

Estimación de la media con una muestra: Al seleccionar una muestra de una población normal, o si n es grande, se establece un intervalo de confianza para μ considerando la distribución muestral de \bar{X} . Esta estimación de la media de la población se realiza con el teorema del límite central. Distribución muestral de \bar{X} :

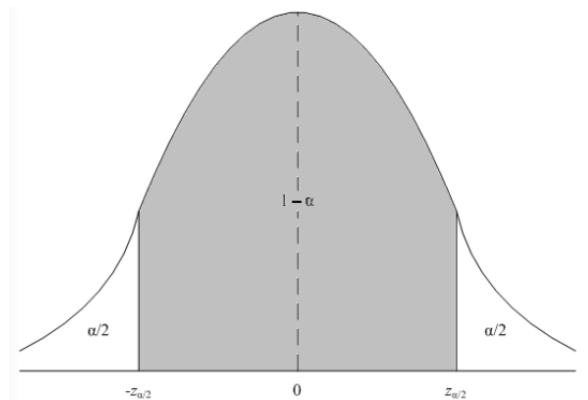
Media:

$$\mu_{\bar{X}} = \mu$$

Desviación Estándar:

$$\sigma_{\bar{X}} = \sigma / \sqrt{n}$$

- Se forma una distribución aproximadamente normal

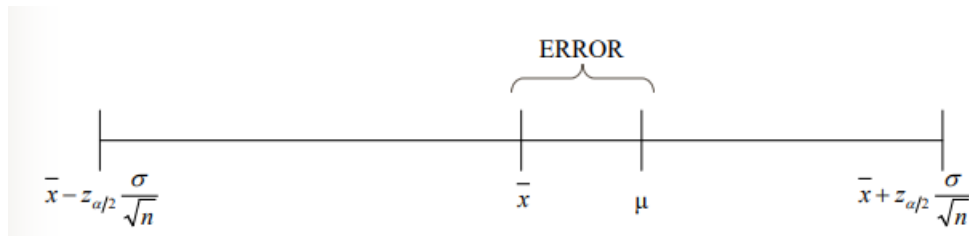


Intervalo de confianza de μ con σ conocida: Para calcular el intervalo de confianza de $(1 - \alpha)100\%$ para μ , de una muestra aleatoria de tamaño n de una población con una **varianza σ^2 conocida**, se necesita la media \bar{X} de esta muestra. Esto está definido por:

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

- **Nota:** Si se cuenta con muestras de tamaño $n \geq 30$, se garantizan buenos resultados, sin importar la forma de las poblaciones. En el caso de que las muestras sean más pequeñas, y provengan de poblaciones no normales, el grado de confianza no necesariamente será preciso.

En los casos en los que μ es el valor central del intervalo, entonces \bar{X} estima μ sin error, pero la mayoría de las veces \bar{x} no es exactamente igual a μ por lo que la estimación puntual contará con un **error**.



Teoremas:

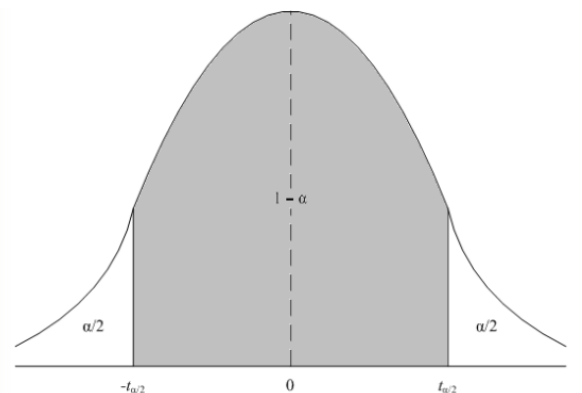
- Al utilizar \bar{X} como una estimación de μ , se tiene una confianza de $(1 - \alpha)100\%$ de que el error nunca excederá:

$$z_{\alpha/2} \sigma / \sqrt{n}.$$

- El error nunca superará este valor cuando el tamaño de la muestra es:

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

En el caso de que no se conozca la varianza de la población, se debe utilizar la distribución t para estimar la media de la población:



Intervalo de Confianza de μ con σ desconocida: Para calcular el intervalo de confianza de $(1 - \alpha)100\%$ para μ , de una muestra aleatoria de tamaño n de una población normal con una **varianza σ^2 desconocida**, se debe de contar con la media \bar{X} y la desviación estándar s . Esto está definido por:

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

Teoremas:

- Al utilizar \bar{X} como una estimación de μ , se tiene una confianza de $(1 - \alpha)100\%$ de que el error nunca excederá:

$$t_{\alpha/2} s / \sqrt{n}.$$

- El error nunca superará este valor cuando el tamaño de la muestra es:

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

Intervalo de confianza de muestra grande: Cuando σ se desconoce y $n \geq 30$, se puede reemplazar a σ incluso cuando no se puede garantizar la normalidad de la muestra:

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

- Entre más grande sea el tamaño de la muestra, mejor será este resultado.

Límites de Confianza Unilaterales: Existen ciertas situaciones en donde sólo se requiere de un límite, en vez de los dos (inferior y superior). Por ejemplo: Si se desea examinar el escenario del peor caso, solo se necesita del límite inferior, o si al medir una variable no se desea un valor grande de la media, se utiliza solamente el límite superior.

- Los límites de Confianza unilaterales se desarrollan de la misma forma que los bilaterales: Para calcular los límites de confianza unilaterales de $(1 - \alpha)100\%$ para μ , se debe de tener una muestra aleatoria de tamaño n de una población con una **varianza σ^2 conocida**, además de la media \bar{X} de esta muestra. Esto está definido por:

Límite inferior:

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

Límite superior:

$$\bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

- Igualmente, para calcular los límites de confianza unilaterales de $(1 - \alpha)100\%$ para μ , de una muestra aleatoria de tamaño n de una población con una **varianza σ^2 desconocida**, se debe de contar con la media \bar{X} . Esto está definido por:

Límite Inferior:

$$\bar{x} - t_{\alpha} \frac{s}{\sqrt{n}}$$

Límite superior:

$$\bar{x} + t_{\alpha} \frac{s}{\sqrt{n}}$$

Error Estándar de la Estimación Puntual: En las estimaciones, la desviación estándar de \bar{X} corresponde al error estándar de \bar{X} . En los casos donde σ se desconoce y la muestra es sobre una distribución normal, s reemplaza a σ , y además se incluye al error estándar estimado en la fórmula. El ancho del intervalo de confianza de μ depende de la calidad de la estimación puntual y consecuentemente de su error estándar.

Límites de confianza de μ :

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

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

- Colocando el error estándar o el error estándar estimado en la fórmula respectivamente.

Límites de Predicción: A veces se necesita predecir los posibles valores de una observación futura. Para realizar esta predicción, se necesita la variación de la media y de la observación futura. Esta predicción se realiza por medio de la construcción de un intervalo de predicción.

- En una distribución normal de mediciones con una **media μ desconocida** y una **varianza σ^2 conocida**, el intervalo de predicción de $(1 - \alpha)100\%$ de una observación futura denominada X_0 corresponde a:

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

- En una distribución normal de mediciones con una **media μ desconocida** y una **varianza σ^2 desconocida**, el intervalo de predicción de $(1 - \alpha)100\%$ de una observación futura denominada X_0 corresponde a:

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

- En este caso se utiliza la distribución t de Student.

Intervalos de predicción unilaterales: En los casos donde se estudian observaciones futuras grandes se utilizan los límites de predicción superiores, y en los casos donde se estudian observaciones futuras pequeñas se utilizan los límites de predicción inferiores:

$$\begin{array}{l} \sigma \text{ conocida} \quad \left\{ \begin{array}{l} \text{límite inferior } \bar{x} - z_{\alpha} \sigma \sqrt{1 + 1/n} \\ \text{límite superior } \bar{x} + z_{\alpha} \sigma \sqrt{1 + 1/n} \end{array} \right. \\ \sigma \text{ desconocida} \quad \left\{ \begin{array}{l} \text{límite inferior } \bar{x} - t_{\alpha} s \sqrt{1 + 1/n} \\ \text{límite superior } \bar{x} + t_{\alpha} s \sqrt{1 + 1/n} \end{array} \right. \end{array}$$

Límites de Tolerancia: En las ocasiones en las que se desea averiguar dónde caen la mayoría de los valores de la población, para conocer el desempeño a largo plazo en vez de la siguiente observación, se puede determinar un intervalo de confianza sobre una proporción fija de las mediciones. Esto se denomina un intervalo de tolerancia, y cubre exactamente el 95% de las observaciones.

- En circunstancias comunes, rara vez se conocen los valores de μ y σ , por lo que los límites de tolerancia bilaterales están dados por:

$$\bar{x} \pm ks$$

Límites de tolerancia unilaterales:

$$\text{límite inferior } \bar{x} - ks$$

$$\text{límite superior } \bar{x} + ks$$

- **Nota:** En los dos casos, k se calcula de manera que se pueda asegurar con una confianza de $(1 - \gamma)100\%$ que los límites incluyen al menos la proporción $(1 - \alpha)$ de las medidas. (ver tabla).

Usos del Límite de Confianza, Límite de Predicción y Límite de Tolerancia:

- **Intervalo de Confianza:** Este se calcula sobre la media cuando se interesa estimar la media de la población, debido a que produce los límites apropiados para esta estimación.
- **Intervalo de Tolerancia:** Sirve para averiguar dónde estará la mayor parte de los valores de la población, u observaciones individuales.
- **Intervalo de Predicción:** Utilizada para determinar un límite de un solo valor, o la ubicación de una sola observación, en vez de la mayoría de la población.

Estimación de la Diferencia entre Medias de Dos Muestras: Al tener dos poblaciones normales, o, si n_1 y n_2 son lo suficientemente grandes, se puede definir un intervalo de confianza para $\mu_1 - \mu_2$, considerando la distribución muestral de $\bar{X}_1 - \bar{X}_2$. La estimación de la diferencia de estas dos medias de dos poblaciones se realiza con el teorema del límite central. Distribuciones muestrales de $\bar{X}_1 - \bar{X}_2$:

Media:

$$\mu_1 - \mu_2$$

Desviación Estándar:

$$\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}.$$

Intervalo de Confianza de $\mu_1 - \mu_2$ con σ_1 y σ_2 conocidos: Para calcular el intervalo de confianza de $(1 - \alpha)100\%$ para $\mu_1 - \mu_2$, cuando se tiene unas muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones con **varianzas σ_1^2 y σ_2^2 conocidas**, se necesitan las medias \bar{X}_1 y \bar{X}_2 de estas muestras. Esto está definido por:

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

- **Nota:** Si se cuenta con muestras de tamaño $n \geq 30$, se garantizan buenos resultados, sin importar la forma de las poblaciones. En el caso de que las muestras sean más pequeñas, y provengan de poblaciones no normales, el grado de confianza no necesariamente será preciso.

Intervalo de Confianza de $\mu_1 - \mu_2$ con σ_1 y σ_2 iguales pero desconocidos: Para calcular el intervalo de confianza de $(1 - \alpha)100\%$ para $\mu_1 - \mu_2$, cuando se tiene unas muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones aproximadamente normales con **varianzas iguales pero desconocidas**, se necesitan las medias \bar{X}_1 y \bar{X}_2 de estas muestras. Esto está definido por:

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

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

Intervalo de Confianza de $\mu_1 - \mu_2$ con σ_1 y σ_2 diferentes pero desconocidos: Para calcular el intervalo de confianza de $(1 - \alpha)100\%$ para $\mu_1 - \mu_2$, cuando se tiene unas muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones aproximadamente normales con **varianzas diferentes y desconocidas**, se necesitan las medias \bar{X}_1 y \bar{X}_2 de estas muestras. Esto está definido por:

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

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

Comparación de dos medias:

- Si el intervalo de confianza de la diferencia $\mu_1 - \mu_2$ es **positivo**, entonces $\mu_1 > \mu_2$ con poco grado de error.
- Si el intervalo de confianza de la diferencia $\mu_1 - \mu_2$ es **negativo**, entonces $\mu_1 < \mu_2$ con poco grado de error.
- Si el intervalo de confianza de la diferencia $\mu_1 - \mu_2$ es igual a 0, entonces $\mu_1 = \mu_2$ con poco grado de error.

Error Estándar de la Estimación Puntual con dos Muestras: La desviación estándar de $\bar{X}_1 - \bar{X}_2$ corresponde al error estándar de $\bar{X}_1 - \bar{X}_2$. En los casos en los que no se conocen σ_1 y σ_2 y el muestreo se da sobre una distribución normal, s_1 y s_2 reemplazan a σ_1 y σ_2 , y además se incluye al error estándar estimado en la fórmula. El ancho del intervalo de confianza de μ_1 y μ_2 depende de la calidad de la estimación puntual y consecuentemente de su error estándar.

Límites de Confianza de $\mu_1 - \mu_2$:

$$\begin{aligned} (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} s.e.(\bar{x}_1 - \bar{x}_2), \quad \sigma_1 \text{ y } \sigma_2 \text{ son conocidas} \\ (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} &= (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \hat{s}.e.(\bar{x}_1 - \bar{x}_2), \quad \sigma_1 \text{ y } \sigma_2 \text{ son diferentes y desconocidas} \\ (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} &= (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \hat{s}.e.(\bar{x}_1 - \bar{x}_2), \quad \sigma_1 \text{ y } \sigma_2 \text{ son iguales y desconocidas} \end{aligned}$$

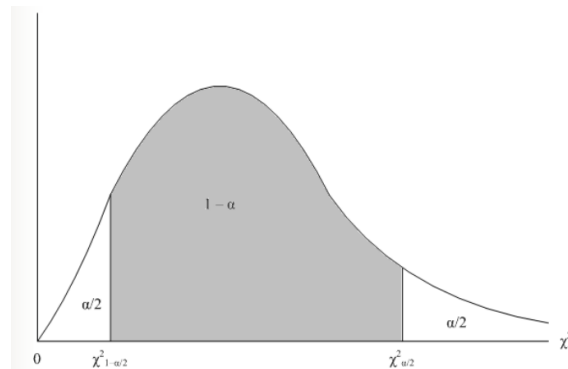
- Se coloca el error estándar o el error estándar estimado en la fórmula dependiendo de si se conocen σ_1 y σ_2 o no.

Estimación de la Varianza con una Muestra: Al extraer una muestra n de una población normal con varianza σ^2 y calcular la varianza muestral s^2 se obtiene una estadística denominada estimador de σ^2 , representada por S^2 .

- Mediante el S^2 , se puede realizar una estimación por intervalos de σ^2 :

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

- Donde χ^2 tiene una distribución chi-cuadrada con $\nu = n - 1$ grados de libertad cuando las muestras provienen de una población normal.



Intervalo de confianza para σ^2 : Para calcular el intervalo de confianza de $(1 - \alpha)100\%$ para σ^2 , de una muestra aleatoria de tamaño n de una población normal, se necesita la varianza s^2 de esta muestra. Esto está definido por:

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

Intervalo de confianza para σ : Para calcular el intervalo de confianza de $(1 - \alpha)100\%$ para σ , se necesita la raíz cuadrada de cada extremo del intervalo de confianza para σ^2 . Esto está definido por:

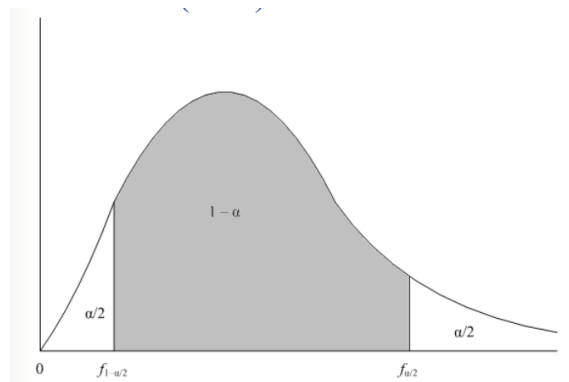
$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}$$

Estimación de la Razón de Dos Varianzas con dos Muestras: La estimación puntual de la razón de dos varianzas σ_1^2 / σ_2^2 se obtiene de la razón s_1^2 / s_2^2 de las **varianzas muestrales**, por lo que esta razón se denomina estimador de σ_1^2 / σ_2^2 .

- En el caso en donde σ_1^2 y σ_2^2 son varianzas de poblaciones normales, se puede calcular una estimación por intervalos de σ_1^2 / σ_2^2 mediante:

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

- En donde F cuenta con una distribución f con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad.



Intervalos de confianza para σ_1^2 / σ_2^2 : Para calcular el intervalo de confianza de $(1-\alpha)100\%$ para σ_1^2 / σ_2^2 , de unas muestras independientes de tamaño n_1 y n_2 de unas poblaciones normales, se necesita las varianzas s_1^2 y s_2^2 de estas muestras para calcular la razón entre ambas. Esto está definido por:

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

Intervalos de confianza para σ_1 / σ_2 : Para calcular el intervalo de confianza de $(1-\alpha)100\%$ para σ_1 / σ_2 , se necesita la raíz cuadrada de cada extremo del intervalo de confianza para σ_1^2 / σ_2^2 . Esto está definido por:

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

Comparación de dos varianzas:

- Si el intervalo de confianza de la razón σ_1^2 / σ_2^2 puede ser igual a 1, entonces $\sigma_1^2 = \sigma_2^2$ con poco grado de error.
- Si el intervalo de confianza de la razón σ_1^2 / σ_2^2 no puede ser igual a 1, entonces $\sigma_1^2 \neq \sigma_2^2$ con poco grado de error.

Comparación de dos desviaciones estándar:

- Si el intervalo de confianza de la razón σ_1 / σ_2 puede ser igual a 1, entonces $\sigma_1 = \sigma_2$ con poco grado de error.
- Si el intervalo de confianza de la razón σ_1 / σ_2 no puede ser igual a 1, entonces $\sigma_1 \neq \sigma_2$ con poco grado de error.

D. Distribución Z y t, contraste de hipótesis, inferencia de varianzas, distribuciones χ^2 y F

Distribución Z

- Si se tiene X_1, X_2, \dots, X_n y estados son variables aleatorias independientes que tienen distribuciones normales con medias μ y varianzas σ^2 , entonces se dice que la siguiente variable aleatoria tiene una distribución normal, con una media y una varianza que se muestran a continuación:

$$Y = a_1X_1 + a_2X_2 + \dots + a_nX_n$$

$$\mu_Y = a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n$$

$$\sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2$$

- De acuerdo con el teorema que indica la propiedad de reproducción de la distribución normal, se infiere que:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

tiene distribución normal con las siguientes medias y varianzas respectivamente:

$$\mu_{\bar{X}} = \frac{\mu + \mu + \dots + \mu}{n} = \mu \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- Si se tiene que \bar{X} contiene el valor de la media de una muestra que es aleatoria y está a su vez tiene un tamaño de n elementos, tomada de una población que presenta una media con valor μ y una varianza con valor σ^2 :

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

Es una distribución normal estándar $n(z;0,1)$ en cuanto n tiende hacia el infinito.

A esto anterior se le menciona con el nombre de Teorema del Límite Central.

- La aproximación normal para \bar{X} será eficaz si se toma en cuenta lo siguiente:
 - Si n toma un valor mayor o igual a 30, la aproximación será buena sin importar qué forma tenga la población.
 - Si n toma un valor menor a 30, la aproximación será buena si la población se parece a una distribución normal.
 - Si en general se sabe que la distribución de la población es normal.
- El teorema anterior es importante a la hora de determinar valores razonables con respecto a la media de una población, e incluso para pruebas de hipótesis, estimaciones, controles de calidad y más.

Distribución t

- En muchas ocasiones se conoce la media de la población, pero no se tiene un conocimiento razonable de σ .
- Comúnmente una estimación de σ la debe dar la información muestral que genera el promedio muestral \bar{x} .

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

- S es el análogo de la muestra para σ .
- Si la muestra es pequeña los valores de S^2 fluctúan entre muestras y la distribución T se desvía de la distribución normal estándar.
- Sin embargo, si la muestra es suficientemente grande, la distribución no cambia considerablemente de la normal.
-
- Si Z es una variable aleatoria normal estándar, V una variable aleatoria ji cuadrado con cantidad v de grados de libertad, entonces:

$$T = \frac{Z}{\sqrt{V/v}}$$

Está dada por:

$$h(t) = \frac{\Gamma[(v+1)/2]}{\Gamma[v/2]\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2} \quad -\infty < t < +\infty$$

- Se le conoce como distribución t con cantidad v de grados de libertad, v toma el valor de n-1 si la muestra tiene como tamaño a n.
- Si se tienen $X_1 \dots X_n$ variables aleatorias independientes, normales con media μ y desviación estándar σ .

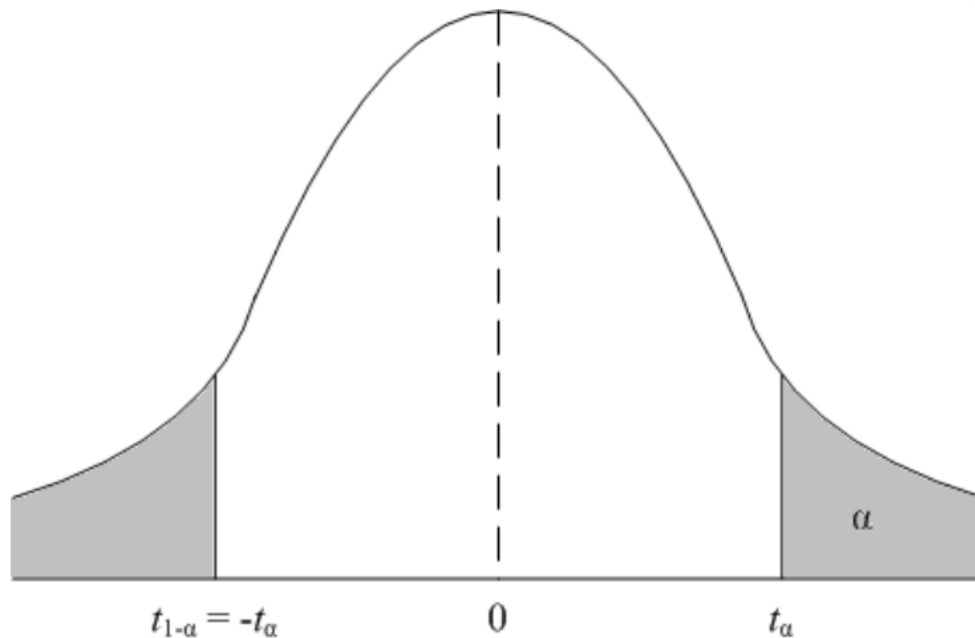
$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n} \quad S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$$

Entonces la variable aleatoria

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

tiene una distribución t con n-1 grados de libertad.

- La distribución t es similar a la distribución Z, esto se debe a que ambas son simétricas alrededor de una media de cero y las 2 distribuciones tienen una forma física de campana.
- Sin embargo, la distribución t es más variable que la distribución normal estándar, ya que los valores de T dependen de los cambios que le suceden a \bar{X} y S^2 , mientras que los valores de Z dependen únicamente de \bar{X} entre muestras.
- Usualmente se le representa con t_α al valor t por encima del que hay un área igual a α .
- Visualización de la distribución t:



- Si un valor se encuentra por debajo de $-t_{0.025}$ o por encima de $t_{0.025}$, se cree que el valor no es común e incluso se puede decir que la suposición hecha de μ
- Si un valor t cae por debajo de $-t_{0.01}$ o por encima de $t_{0.01}$ proporciona buen fundamento por el cual pensar que el valor supuesto de μ es muy poco probable.
- Esta distribución se usa de manera frecuente en problemas que se desean inferencia acerca de la media poblacional o comparar muestras.
- Usar esta distribución y considerar el tamaño de la muestra no están relacionados con el teorema del límite central.
- Usar la distribución normal estándar en vez de T para una cantidad n mayor o igual que 30 implica que S estima de buena manera a σ .

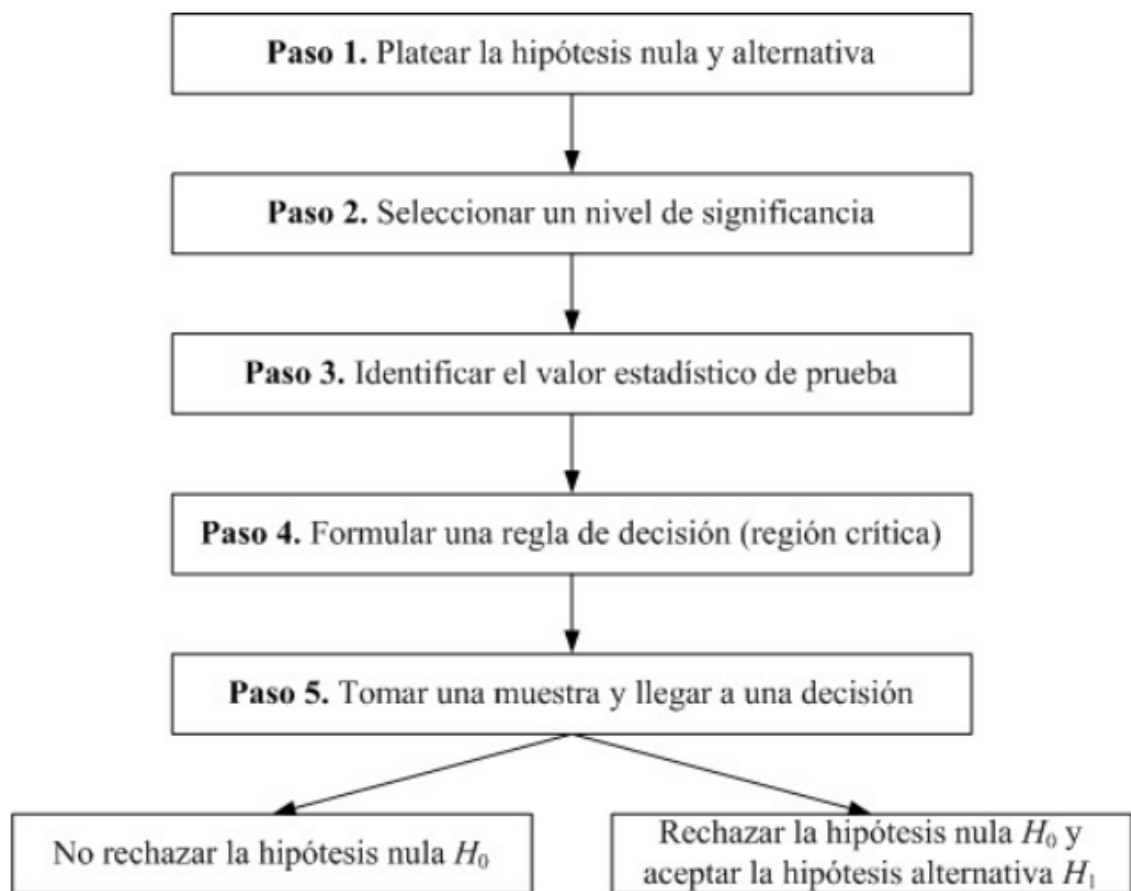
Contraste de hipótesis

- El contraste de hipótesis estadística es una regla o procedimiento que conduce a la resolución de una decisión en la cual se acepta o se rechaza una hipótesis propuesta, con base a un resultado o varios resultados.
- El procedimiento de contraste de hipótesis depende de la forma en la que se utiliza la información contenida en una muestra aleatoria de la población.
 - Si la información es consistente con la hipótesis se toma la decisión que sí es verdadera. Si no lo es, entonces se dice que es falsa.
- El estadístico de prueba es el valor obtenido a partir de la información muestral, se utiliza para determinar si se rechaza o no la hipótesis.
- Se le llama C a la región crítica de una prueba si esta nos lleva a rechazar a la hipótesis nula H_0 cuando la muestra cae en un valor de la región C .

- El valor crítico es un punto que divide la región aceptada y la región de rechazo de la hipótesis nula H_0 .
- El nivel de significancia es la posibilidad de rechazar la hipótesis nula cuando es verdadera.
- El error tipo I es rechazar la hipótesis nula cuando es verdadera. La probabilidad de caer en un error de tipo I es el nivel de significancia.
- El error tipo II es aceptar la hipótesis nula cuando era falsa.
- Es imposible caer en error tipo 2 si no se tiene una hipótesis alternativa.

		Situaciones Posibles	
		H_0 verdadera	H_0 falsa
Decisiones Posibles	No rechazar H_0	Decisión correcta	Error tipo II
	Rechazar H_0	Error tipo I	Decisión correcta

- Rechazar la hipótesis nula es importante, esta decisión se basa en una muestra y no en una población, esto genera que se pueda cometer error tipo I.



- Las probabilidades de los errores de tipo I y II están relacionadas, si una disminuye otra aumenta.
- Si se aumenta la muestra se disminuye α y β a la misma vez.

- Si la hipótesis nula es falsa, β es un máximo en el momento que el valor real de un parámetro se acerque hacia el valor hipotético.
- El poder de la potencia es la probabilidad de que se rechace la hipótesis nula cuando ésta es falsa y correctamente debe ser rechazada, se desea que el valor de ésta sea alto.
 - Si se utiliza un poder muy bajo, la posibilidad de que ocurra un error aumenta mucho, por lo que genera resultados en los cuales se puede confiar poco.

Inferencia de Varianzas

- Si se tiene S^2 como la varianza de una muestra aleatoria de tamaño n tomada de una población normal que tiene una varianza de σ^2 , entonces se puede decir que

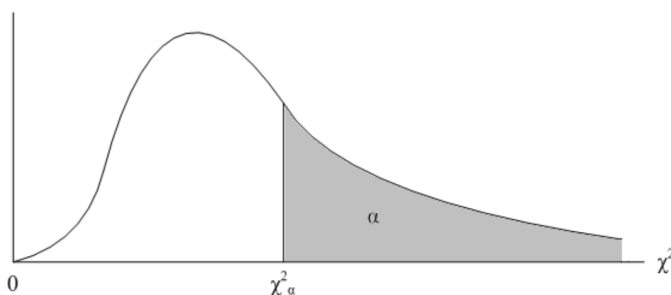
$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

tiene distribución chi cuadrado con $v=n-1$ cantidad de grados de libertad.

Distribución Chi-Cuadrado

- Si se asume que S^2 toma el valor que tiene la varianza de la muestra aleatoria, tomada de una población normal que tiene como valor de varianza σ^2 , se puede decir que la siguiente fórmula cuenta con una distribución chi cuadrado y cuenta también con una cantidad de $n-1$ grados de libertad:

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



- Mucha de la distribución (95%) chi cuadrado está entre $X^2_{0.975}$ y $X^2_{0.025}$.
- Solamente si σ^2 es muy pequeño, se puede decir que un valor que cae del lado derecho de $X^2_{0.025}$ es probable que ocurra.
- Solamente si σ^2 es muy grande, se puede decir que un valor que cae del lado izquierdo de $X^2_{0.975}$ es probable que ocurra.

Distribución F

- Esta distribución es muy aplicada a la hora de comparar varianzas de una o incluso varias muestras.
- La estadística F se puede ilustrar como la siguiente fórmula:

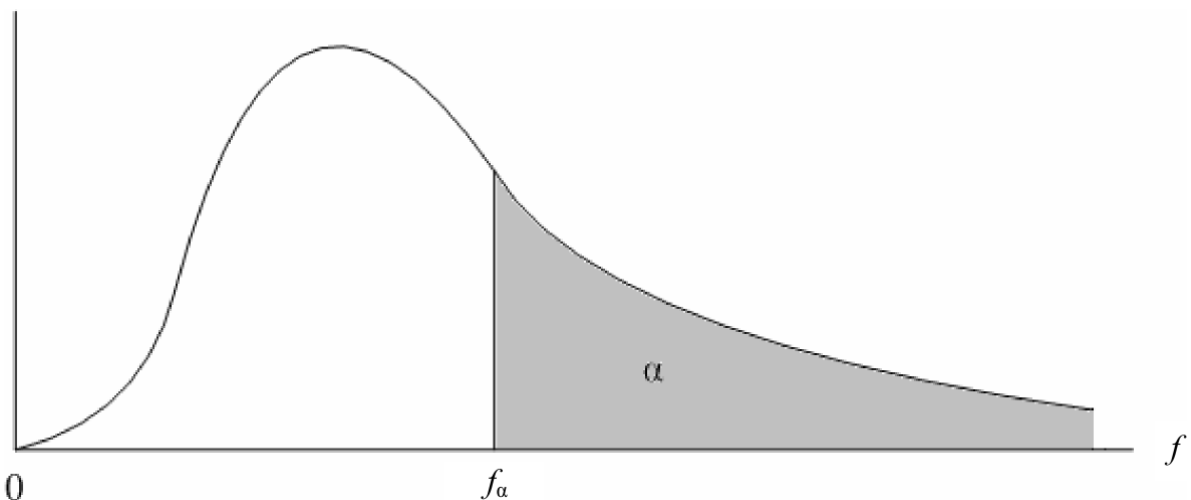
$$F = \frac{U/v_1}{V/v_2}$$

U y V son variables aleatorias independientes que presentan distribuciones ji cuadradas cada una con una cantidad v grados de libertad. La distribución de variable aleatoria F está dada por:

$$h(f) = \begin{cases} \frac{\Gamma[(v_1 + v_2)/2] (v_1/v_2)^{v_1/2}}{\Gamma(v_1/2) \Gamma(v_2/2)} * \frac{f^{v_1/2-1}}{(1 + v_1 f/v_2)^{(v_1+v_2)/2}} & 0 < f < +\infty \\ 0 & \text{otro caso} \end{cases}$$

y se le conoce como distribución F con v_1 y v_2 cantidad de grados de libertad.

- La curva de esta distribución depende del orden en que establezcan v_1 y v_2 y los valores que se le den a estas.
- Curva de la distribución:



- Si se escribe $f_{\alpha}(v_1, v_2)$, teniendo v_1, v_2 cantidad de grados de libertad para f_{α} , se obtiene lo siguiente:

$$f_{1-\alpha}(v_1, v_2) = \frac{1}{f_{\alpha}(v_2, v_1)}$$

- Si se tiene que S_1^2 y S_2^2 son los valores de las varianzas de muestras aleatorias independientes de tamaño n_1 y n_2 que fueron anteriormente tomadas de poblaciones normales con varianzas σ_1^2 y σ_2^2 , respectivamente se puede mostrar la siguiente fórmula:

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

y se tendría una distribución F con los siguientes grados de libertad: $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$.

- Esta distribución es normalmente utilizada en situaciones dónde se tienen 2 muestras y de estas se desea obtener inferencias acerca de los valores de las varianzas de una población, sin embargo también es aplicada a problemas que contienen varianzas.
- El nombre de esta distribución es: distribución de razón de varianzas.