Universidad del País Vasco
Euskal Herriko Unibertsitatea

Material de estudio

OCW 2019: *Curso práctico para el análisis e inferencia estadística con Mathematica*

Tema 7. Intervalos de confianza

Equipo docente del curso
Arrospide Zabala, Eneko
Martín Yagüe, Luis
Unzueta Inchaurbe, Aitziber
Soto Merino, Juan Carlos
Durana Apaolaza, Gaizka
Bikandi Irazabal, Iñaki

Departamento de Matemática Aplicada
Escuela de Ingeniería de Bilbao, Edificio II-I

TEMA 7. INTERVALOS DE CONFIANZA

Introducción

Definición

Un estimador por intervalo es una regla ó fórmula que, usando la información muestral, permite hallar dos valores que definen un intervalo utilizado como estimación del parámetro poblacional con cierto grado de confianza.

Sea $\{x_1, x_2, \ldots, x_n\}$ una muestra aleatoria simple con n observaciones de una variable aleatoria, X, cuya distribución depende de un parámetro desconocido θ.

Un intervalo de confianza para θ con un nivel de significación α, $I_\theta^{1-\alpha}(x_1, x_2, \ldots, x_n)$, es un intervalo real que depende de la muestra, pero que no depende de θ, tal que:

$$\boxed{P[\theta \in I_\theta^{1-\alpha}(x_1, x_2, \ldots, x_n)] = 1-\alpha}$$

Se denomina nivel de confianza al valor $1-\alpha$.

Procedimiento

El objetivo de un intervalo de confianza es proporcionar, en base a los datos de la muestra, una región en la que se tenga un determinado nivel de confianza de que se encuentre el parámetro.

Como en el caso de los estimadores puntuales (aproximaciones de los parámetros mediante un único valor), el intervalo de confianza es aleatorio ya que depende de los valores de la muestra.

Para obtener un intervalo de confianza se debe:

- identificar un estadístico que dependa de los valores muestrales y del parámetro que se está estimando
- seleccionar un nivel de confianza
- aplicar el estimador por intervalo correspondiente

Existe la posibilidad de que el verdadero parámetro poblacional, θ, no tome un valor perteneciente al intervalo de confianza, algo que ocurre con probabilidad α.

Estimación de la media poblacional

Supuestos adicionales: población normal y σ conocida

Estimación puntual: $\overline{x}$

Intervalo de confianza:

$$\boxed{I_\mu^{1-\alpha} = \left(\overline{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de la muestra obtenida o utilizando las siguientes funciones:

- `MeanCI[`*list*`,ConfidenceLevel→1-`α`,KnownVariance→`σ^2`]`. Da un intervalo de confianza para la media poblacional estimada a partir de la lista *list* formada por n elementos. Puede indicarse la varianza si es conocida; en ese caso, se basa en una distribución normal. El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.
- `NormalCI[`$\overline{x}$`,`$\sigma/\sqrt{n}$`,ConfidenceLevel→1-`α`]`. Proporciona un intervalo de confianza para la media poblacional cuando la varianza poblacional es conocida. Se basa en una distribución normal estimando los parámetros a partir de la media muestral ($\overline{x}$) y el error estándar ($\sigma/\sqrt{n}$). El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.

Supuestos adicionales: población normal y σ desconocida

Estimación puntual: $\overline{x}$

Intervalo de confianza:

$$\boxed{I_{\mu}^{1-\alpha}=\left(\overline{x}-t_{n-1;\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}},\ \overline{x}+t_{n-1;\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}}\right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de la muestra obtenida o utilizando las siguientes funciones:

- `MeanCI[`*list*`,ConfidenceLevel→1-`α`,KnownVariance→None]`. Da un intervalo de confianza para la media poblacional estimada a partir de la lista *list* formada por n elementos. Al ser la varianza desconocida, se basa en una distribución t de Student con $\nu=n-1$ grados de libertad. El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.
- `StudentTCI[`$\overline{x}$`,`$S/\sqrt{n}$`,`*df*`,ConfidenceLevel→1-`α`]`. Proporciona un intervalo de confianza para la media poblacional cuando la varianza poblacional es desconocida. Se basa en una distribución t de Student con *df* grados de libertad. El intervalo se estima a partir de la media muestral ($\overline{x}$) y su error estándar ($S/\sqrt{n}$). El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.

Supuestos adicionales: cualquier población y σ conocida ($n>30$)

Estimación puntual: $\overline{x}$

Intervalo de confianza:

$$\boxed{I_{\mu}^{1-\alpha}=\left(\overline{x}-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}},\ \overline{x}+z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de la muestra obtenida o utilizando las funciones `MeanCI` y `NormalCI`.

Supuestos adicionales: cualquier población y σ desconocida ($n>100$)

Estimación puntual: $\overline{x}$

Intervalo de confianza:

$$\boxed{I_{\mu}^{1-\alpha}=\left(\overline{x}-z_{\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}},\ \overline{x}+z_{\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}}\right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de la muestra obtenida o utilizando las funciones `MeanCI` y `NormalCI`.

Ejemplo

El peso de los niños de seis meses sigue una distribución normal de media μ y desviación típica σ. Para analizar el peso medio de los niños en una región de un determinado país se toma la siguiente muestra: $\{6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950\}$. Se pide obtener un intervalo al 90% de confianza para la media de los pesos siendo: a) $\sigma = 1.2\,kg$ y b) σ desconocida

- carga del paquete necesario para las funciones de intervalos de confianza

```
Needs["HypothesisTesting`"] (* el paquete queda cargado para toda la sesión *)
```

- definición de la muestra

```
lista = {6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950};
```

- definición y cálculo de valores

```
α = 0.10; n = Length[lista];
med = Mean[lista]; S = StandardDeviation[lista];
```

- a) $\sigma = 1.2\,kg$
  - definición de σ

```
σ = 1.2;
```

  - usando MeanCI

```
MeanCI[lista, ConfidenceLevel → 1 - α, KnownVariance → σ^2]
{5.44123, 6.63149}
```

  - usando NormalCI

```
NormalCI[Mean[lista], σ / Sqrt[n], ConfidenceLevel → 1 - α]
{5.44123, 6.63149}
```

  - simulando el estimador

```
{Mean[lista] - Quantile[NormalDistribution[0, 1], q = 1 - α / 2] * σ / Sqrt[n],
 Mean[lista] + Quantile[NormalDistribution[0, 1], q = 1 - α / 2] * σ / Sqrt[n]}
{5.44123, 6.63149}
```

  - solución: $\boxed{I_{\mu}^{0.90} = (5.44123, 6.63149)}$

- b) σ desconocida
  - usando MeanCI

```
MeanCI[lista, ConfidenceLevel → 1 - α, KnownVariance → None]
{5.86922, 6.2035}
```

  - usando StudentTCI

```
StudentTCI[med, S / Sqrt[n], n - 1, ConfidenceLevel → 1 - α]
{5.86922, 6.2035}
```

  - simulando el estimador

```
{med - Quantile[StudentTDistribution[n - 1], q = 1 - α / 2] * S / Sqrt[n],
 med + Quantile[StudentTDistribution[n - 1], q = 1 - α / 2] * S / Sqrt[n]}
{5.86922, 6.2035}
```

- solución: $\boxed{I_{\mu}^{0.90} = (5.86922,\ 6.2035)}$

Estimación de la diferencia de dos medias poblacionales

Muestras independientes

Sean dos poblaciones diferentes de medias μ_1 y μ_2 y desviaciones típicas σ_1 y σ_2 , respectivamente.

Se consideran dos muestras independientes extraídas aleatoriamente de esas poblaciones. Los tamaños de ambas muestras se denotan n_1 y n_2, respectivamente.

Dos muestras son independientes si la selección de los datos de una población no está relacionada con la de los datos de la otra.

Se va a utilizar la información suministrada por las dos muestras para estimar la diferencia de las medias poblacionales: $\mu_1 - \mu_2$.

Mathematica dispone de la función:

- **MeanDifferenceCI[*list1*,*list2*,ConfidenceLevel→1-α,EqualVariances→False,KnownVariance→None]**. Proporciona un intervalo de confianza para la diferencia de medias poblacionales estimada a partir de las listas *list1* y *list2* formadas por n_1 y n_2 elementos, respectivamente. Necesita la carga del paquete `HypothesisTesting`` .
  - **ConfidenceLevel**. El nivel de confianza por defecto es 0.95.
  - **EqualVariances**. Puede asumirse el caso de varianzas desconocidas pero iguales asignando el valor lógico `True` en este argumento (por defecto, `False`); en este caso, se basa en una distribución t de Student con $\nu = n_1 + n_2 - 1$ grados de libertad.
  - **KnownVariance**. Las varianzas de las poblaciones son conocidas y se introducen como elementos de una lista (por defecto, `None`); en este caso, se basa en una distribución normal.

Muestras dependientes (datos pareados)

Si las muestras se seleccionan de manera que cada medida en una de ellas pueda asociarse naturalmente con una medida en la otra muestra se llaman muestras dependientes.

Si dos muestras son dependientes entonces, necesariamente, tienen el mismo tamaño; los datos van emparejados por individuos.

Para estimar la diferencia de las medias poblacionales $(\mu_1 - \mu_2)$ se obtiene una nueva muestra como la diferencia entre los valores correspondientes a un mismo individuo en las dos muestras dadas. Así, el problema se reduce a la estimación de una media.

Supuestos adicionales: poblaciones normales y σ_1 y σ_2 conocidas

Estimación puntual: $\overline{x}_1 - \overline{x}_2$

Intervalo de confianza:

$$\boxed{I_{\mu_1-\mu_2}^{1-\alpha} = \left((\overline{x}_1 - \overline{x}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\ ,\ (\overline{x}_1 - \overline{x}_2) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de las muestras dadas o utilizando la función:

- **MeanDifferenceCI[*list1*,*list2*,ConfidenceLevel→1-α,EqualVariances→False,KnownVariance→$\{\sigma_1^2, \sigma_2^2\}$]**.

Supuestos adicionales: poblaciones normales y σ_1 y σ_2 desconocidas pero iguales

Estimación puntual: $\bar{x}_1 - \bar{x}_2$

Intervalo de confianza:

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left((\bar{x}_1 - \bar{x}_2) - t_{n_1+n_2-1;\frac{\alpha}{2}} \cdot S_P \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \, , \, (\bar{x}_1 - \bar{x}_2) + t_{n_1+n_2-1;\frac{\alpha}{2}} \cdot S_P \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Siendo: $S_P = \sqrt{\dfrac{(n_1-1)\cdot S_1^2 + (n_2-1)\cdot S_2^2}{n_1+n_2-2}}$

El intervalo puede calcularse simulando la fórmula anterior a partir de las muestras dadas o utilizando la función:

- `MeanDifferenceCI[`*list1*,*list2*`,ConfidenceLevel→1-`α`,EqualVariances→True,KnownVariance→None]`.

Supuestos adicionales: poblaciones cualesquiera y σ_1 y σ_2 conocidas (n_1, $n_2 > 15$)

Estimación puntual: $\bar{x}_1 - \bar{x}_2$

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left((\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \, , \, (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

El intervalo puede calcularse simulando la fórmula anterior a partir de las muestras dadas o utilizando la función:

- `MeanDifferenceCI[`*list1*,*list2*`,ConfidenceLevel→1-`α`,EqualVariances→False,KnownVariance→`$\{\sigma_1^2, \sigma_2^2\}$`]`.

Supuestos adicionales: poblaciones cualesquiera y σ_1 y σ_2 desconocidas (n_1, $n_2 > 100$)

Estimación puntual: $\bar{x}_1 - \bar{x}_2$

Intervalo de confianza:

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left((\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \, , \, (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

El intervalo puede calcularse simulando la fórmula anterior a partir de las muestras dadas o utilizando la función:

- `MeanDifferenceCI[`*list1*,*list2*`,ConfidenceLevel→1-`α`,EqualVariances→False,KnownVariance→None]`.

Supuestos adicionales: poblaciones cualesquiera y σ_1 y σ_2 desconocidas y distintas (muestras pequeñas)

Estimación puntual: $\bar{x}_1 - \bar{x}_2$

Intervalo de confianza:

$$I_{\mu_1-\mu_2}^{1-\alpha} = \left((\bar{x}_1 - \bar{x}_2) - t_{\nu;\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \, , \, (\bar{x}_1 - \bar{x}_2) + t_{\nu;\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

Los grados de libertad se calculan con la aproximación de Welch: $\nu = \dfrac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}} - 2$

El intervalo puede calcularse simulando la fórmula anterior a partir de las muestras dadas o utilizando la función:

- `MeanDifferenceCI[`*list1*,*list2*`,ConfidenceLevel→1-`α`,EqualVariances→False,KnownVariance→None]`.

Ejemplo

El peso de los niños de seis meses sigue una distribución normal de media μ y desviación típica σ. Para analizar la diferencia de medias de los pesos de los niños en dos regiones de un país se toman dos muestras independientes: $\{6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950\}$ y $\{6.000, 5.650, 5.750, 5.950, 6.200, 6.250, 5.800, 5.950, 6.350, 6.050\}$. Se quiere obtener un intervalo al 95% de confianza para la diferencia de medias de ambas poblaciones suponiendo: a) $\sigma_1 = 1.2\,kg$, $\sigma_2 = 1.5\,kg$, b) σ_1, σ_2 desconocidas pero iguales y c) σ_1, σ_2 desconocidas y desiguales

- definición de las muestras

```
muestra1 = {6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950};
muestra2 = {6.000, 5.650, 5.750, 5.950, 6.200, 6.250, 5.800, 5.950, 6.350, 6.050};
```

- definición y cálculo de valores

```
α = 0.05; n1 = Length[muestra1]; n2 = Length[muestra2];

m1 = Mean[muestra1]; S1 = StandardDeviation[muestra1];

m2 = Mean[muestra2]; S2 = StandardDeviation[muestra2];
```

- a) $\sigma_1 = 1.2\,kg$, $\sigma_2 = 1.5\,kg$
  - definición de σ_1, σ_2

```
σ1 = 1.2; σ2 = 1.5;
```

  - usando `MeanDifferenceCI`

```
MeanDifferenceCI[muestra1, muestra2, KnownVariance → {σ1^2, σ2^2}]

{-1.12791, 1.21064}
```

  - simulando el estimador

```
{(m1 - m2) - Quantile[NormalDistribution[], q = 1 - α/2] * Sqrt[σ1^2/n1 + σ2^2/n2],
  (m1 - m2) + Quantile[NormalDistribution[], q = 1 - α/2] * Sqrt[σ1^2/n1 + σ2^2/n2]}

{-1.12791, 1.21064}
```

  - solución: $\boxed{I^{0.95}_{\mu_1-\mu_2} = (-1.12791, 1.21064)}$

- b) σ_1, σ_2 desconocidas pero iguales
  - usando `MeanDifferenceCI`

```
MeanDifferenceCI[muestra1, muestra2, EqualVariances → True, KnownVariance → None]

{-0.20623, 0.288958}
```

- simulando el estimador

```
sp = Sqrt[((n1 - 1) * S1^2 + (n2 - 1) * S2^2)/(n1 + n2 - 2)];
{(m1 - m2) - Quantile[StudentTDistribution[n1 + n2 - 2], q = 1 - α / 2] * sp * Sqrt[1 / n1 + 1 / n2],
 (m1 - m2) + Quantile[StudentTDistribution[n1 + n2 - 2], q = 1 - α / 2] * sp * Sqrt[1 / n1 + 1 / n2]}
{-0.20623, 0.288958}
```

- solución: $\boxed{I^{0.95}_{\mu_1-\mu_2} = (-0.20623,\ 0.288958)}$

- c) σ_1, σ_2 desconocidas y desiguales
  - usando **MeanDifferenceCI**

```
MeanDifferenceCI[muestra1, muestra2, EqualVariances → False, KnownVariance → None]
{-0.203245, 0.285973}
```

- simulando el estimador

```
gl = (S1^2/n1 + S2^2/n2)^2/((S1^2/n1)^2/(n1-1) + (S2^2/n2)^2/(n2-1));
{(m1 - m2) - Quantile[StudentTDistribution[gl], q = 1 - α / 2] * Sqrt[S1^2 / n1 + S2^2 / n2],
 (m1 - m2) + Quantile[StudentTDistribution[gl], q = 1 - α / 2] * Sqrt[S1^2 / n1 + S2^2 / n2]}
{-0.203245, 0.285973}
```

- solución: $\boxed{I^{0.95}_{\mu_1-\mu_2} = (-0.203245,\ 0.285973)}$

- **Nota**. En los tres casos, las medias pueden considerarse iguales con un 95% de confianza ya que el valor cero pertenece a todos los intervalos calculados.

Estimación de la varianza poblacional

Supuestos adicionales: población normal y μ desconocida

Estimación puntual: S^2

Intervalo de confianza:

$$\boxed{I^{1-\alpha}_{\sigma^2} = \left(\frac{(n-1)\cdot S^2}{\chi^2_{n-1;\frac{\alpha}{2}}},\ \frac{(n-1)\cdot S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de la muestra obtenida o utilizando las siguientes funciones:

- **VarianceCI[*list*,ConfidenceLevel→1-α]**. Proporciona un intervalo de confianza para la varianza poblacional estimada a partir de la lista *list* formada por n elementos. El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.
- **ChiSquareCI[S^2,*df*,ConfidenceLevel→1-α]**. Calcula un intervalo de confianza para la varianza poblacional cuando la media poblacional es desconocida. Se basa en una distribución *ji*-cuadrado con *df* grados de libertad. El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.

Supuestos adicionales: población normal y μ conocida

Estimación puntual: $\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$

Intervalo de confianza:

$$\boxed{I_{\sigma^2}^{1-\alpha}=\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n;\frac{\alpha}{2}}},\ \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n;1-\frac{\alpha}{2}}}\right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de la muestra obtenida.

Ejemplo

El peso de los niños de seis meses sigue una distribución normal de media μ y desviación típica σ. Para analizar la varianza de los pesos de los niños en una región de un país se toma la siguiente muestra: $\{6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950\}$. Se pide obtener un intervalo al 90% de confianza para la varianza de los pesos siendo: a) $\mu = 6.0\,kg$ y b) μ desconocida

- definición de la muestra

```
muestra1 = {6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950};
```

- definición y cálculo de valores

```
α = 0.10; n1 = Length[muestra1];
m1 = Mean[muestra1]; S1 = StandardDeviation[muestra1];
```

- a) $\mu = 6.0\,kg$
  - definición de μ

```
mu = 6.0;
```

  - simulando el estimador

```
{Total[(muestra1 - mu)^2] / Quantile[ChiSquareDistribution[n1], 1 - α/2],
 Total[(muestra1 - mu)^2] / Quantile[ChiSquareDistribution[n1], α/2]}
{0.0482843, 0.207659}
```

  - solución: $\boxed{I_{\sigma^2}^{0.90}=(0.0482843,\ 0.207659)}$

- b) μ desconocida
  - usando **VarianceCI**

```
VarianceCI[muestra1, ConfidenceLevel → 1 - α]
{0.0510981, 0.237407}
```

  - usando **ChiSquareCI**

```
ChiSquareCI[S1^2, n1 - 1, ConfidenceLevel → 1 - α]
{0.0510981, 0.237407}
```

  - simulando el estimador

$$\boxed{I_{\sigma^2}^{1-\alpha}=\left(\frac{(n-1)\cdot S^2}{\chi^2_{n-1;\frac{\alpha}{2}}},\ \frac{(n-1)\cdot S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right)}$$

```
{(n1 - 1) * S1^2 / Quantile[ChiSquareDistribution[n1 - 1], 1 - α / 2],
 (n1 - 1) * S1^2 / Quantile[ChiSquareDistribution[n1 - 1], α / 2]}
```

```
{0.0510981, 0.237407}
```

- solución: $\boxed{I_{\sigma^2}^{0,90} = (0.0510981,\ 0.237407)}$

Estimación del cociente de dos varianzas poblacionales

Supuestos adicionales: poblaciones normales y μ_1 y μ_2 desconocidas

Estimación puntual: $\dfrac{S_1^2}{S_2^2}$

Intervalo de confianza:

$$\boxed{I_{\frac{\sigma_1^2}{\sigma_2^2}}^{1-\alpha} = \left(\frac{\frac{S_1^2}{S_2^2}}{F_{n_1-1;n_2-1;\frac{\alpha}{2}}},\ \frac{\frac{S_1^2}{S_2^2}}{F_{n_1-1;n_2-1;1-\frac{\alpha}{2}}} \right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de las muestras dadas o utilizando las siguientes funciones:

- **VarianceRatioCI[*list1*,*list2*,ConfidenceLevel→1-α]**. Da un intervalo de confianza para el cociente de dos varianzas poblacionales estimado a partir de las listas *list1* y *list2* formadas por n_1 y n_2 elementos, respectivamente. El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.
- **FRatioCI[S_1^2 / S_2^2,*n*,*m*,ConfidenceLevel→1-α]**. Calcula un intervalo de confianza para el cociente de dos varianzas poblacionales. Se basa en una distribución F de Fisher con n y m grados de libertad en el numerador y el denominador, respectivamente. Los grados de libertad se corresponden con los tamaños muestrales menos una unidad: $n_1 - 1$ y $n_2 - 1$. El nivel de confianza por defecto es 0.95. Necesita la carga del paquete `HypothesisTesting``.

Supuestos adicionales: poblaciones normales y μ_1 y μ_2 conocidas

Estimación puntual: $\dfrac{\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2}{n_1}}{\frac{\sum_{i=1}^{n_2}(y_i-\mu_2)^2}{n_2}}$

Intervalo de confianza:

$$\boxed{I_{\frac{\sigma_1^2}{\sigma_2^2}}^{1-\alpha} = \left(\frac{\frac{n_2 \cdot \sum_{i=1}^{n_1}(x_i-\mu_1)^2}{n_1 \cdot \sum_{i=1}^{n_2}(y_i-\mu_2)^2}}{F_{n_1;n_2;\frac{\alpha}{2}}},\ \frac{\frac{n_2 \cdot \sum_{i=1}^{n_1}(x_i-\mu_1)^2}{n_1 \cdot \sum_{i=1}^{n_2}(y_i-\mu_2)^2}}{F_{n_1;n_2;1-\frac{\alpha}{2}}} \right)}$$

El intervalo puede calcularse simulando la fórmula anterior a partir de la muestra obtenida.

Ejemplo

El peso de los niños de seis meses sigue una distribución normal de media μ y desviación típica σ. Para analizar el cociente de varianzas de los pesos de los niños en dos regiones de un país se toman dos muestras independientes: $\{6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950\}$ y $\{6.000, 5.650, 5.750, 5.950, 6.200, 6.250, 5.800, 5.950, 6.350, 6.050\}$. Se pide obtener un intervalo al 95% de confianza para el cociente de varianzas de ambas poblaciones siendo: a) $\mu_1 = 6.000\,kg$, $\mu_2 = 5.950\,kg$, y b) μ_1, μ_2 desconocidas.

- definición de las muestras

```
muestra1 = {6.100, 5.450, 5.850, 6.550, 6.400, 6.250, 6.000, 5.950, 6.150, 5.750, 5.950};
muestra2 = {6.000, 5.650, 5.750, 5.950, 6.200, 6.250, 5.800, 5.950, 6.350, 6.050};
```

- definición y cálculo de valores

```
α = 0.05; n1 = Length[muestra1]; n2 = Length[muestra2];

m1 = Mean[muestra1]; S1 = StandardDeviation[muestra1];

m2 = Mean[muestra2]; S2 = StandardDeviation[muestra2];
```

- a) $\mu_1 = 6.000\,kg$, $\mu_2 = 5.950\,kg$
  - definición de μ_1, μ_2

```
mu1 = 6.000; mu2 = 5.950;
```

  - simulando el estimador

```
f1 = Quantile[FRatioDistribution[n1, n2], 1 - α / 2];
f2 = Quantile[FRatioDistribution[n1, n2], α / 2];

{(n2 * Total[(muestra1 - mu1) ^ 2] / n1 * Total[(muestra2 - mu2) ^ 2]) / f1,
 (n2 * Total[(muestra1 - mu1) ^ 2] / n1 * Total[(muestra2 - mu2) ^ 2]) / f2}

{0.112523, 1.45394}
```

  - solución: $\boxed{I^{0,90}_{\sigma_1^2/\sigma_2^2} = (0.112523,\ 1.45394)}$

- b) μ_1, μ_2 desconocidas
  - usando **VarianceRatioCI**

```
VarianceRatioCI[muestra1, muestra2, ConfidenceLevel → 1 - α]

{0.464507, 6.95799}
```

  - usando **FRatioCI**

```
FRatioCI[S1^2 / S2^2, n1 - 1, n2 - 1, ConfidenceLevel → 1 - α]

{0.464507, 6.95799}
```

  - simulando el estimador

```
{S1^2 / S2^2 / Quantile[FRatioDistribution[n1 - 1, n2 - 1], 1 - α / 2],
 S1^2 / S2^2 / Quantile[FRatioDistribution[n1 - 1, n2 - 1], α / 2]}

{0.464507, 6.95799}
```

  - solución: $\boxed{I^{0,90}_{\sigma_1^2/\sigma_2^2} = (0.464507,\ 6.95799)}$

- **Nota**. En los dos casos, las varianzas de las dos poblaciones pueden considerarse iguales con un 95% de confianza ya que el valor uno pertenece a todos los intervalos calculados.

Estimación de una proporción poblacional

Supuestos adicionales: población binomial y muestra grande ($n > 100$)

Estimación de la proporción p de éxitos, es decir, la proporción p de individuos de una población que presentan una determinada característica (etiquetada como *éxito*).

Para ello, se toma una muestra aleatoria simple de tamaño, n, y se cuenta el número de éxitos, x .

Estimación puntual: $\hat{p} = \frac{x}{n}$

Intervalo de confianza:

$$\boxed{I_p^{1-\alpha} = \left(\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\ ,\ \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \right)}$$

Ejemplo

Un taller utiliza un modelo de máquina en la fabricación de una determinada pieza para un proceso productivo. Esa pieza debe tener unas medidas que cumplan ciertas especificaciones para resultar válidas. Se quiere determinar la proporción de piezas fabricadas por la máquina que no cumplan con las especificaciones y, por tanto, no sean aceptables para el proceso productivo. Para ello, se recoge una muestra aleatoria simple formada por 285 piezas de las que trece no resultan válidas. Se pide obtener un intervalo al 90% de confianza para la proporción de piezas que no cumplan las especificaciones de medida.

- nivel de significación, número de piezas no válidas (éxitos) en la muestra y tamaño muestral

```
α = 0.10; exitos = 13; n = 285;
```

- proporción éxitos en la muestra

```
p̂ = exitos / n;
```

- simulación el estimador

```
z_(α/2) = Quantile[NormalDistribution[0, 1], 1 - α / 2];
```

```
{p̂ - z_(α/2) * Sqrt[p̂ (1 - p̂) / n], p̂ + z_(α/2) * Sqrt[p̂ (1 - p̂) / n]}
```

```
{0.025285, 0.065943}
```

- solución: $\boxed{I_p^{0.90} = (0.025285,\ 0.065943)}$

Estimación de la diferencia de dos proporciones poblacionales

Supuestos adicionales: poblaciones binomiales y muestras grandes ($n_1, n_2 > 100$)

Estimación de la diferencia entre las proporciones de éxitos de dos poblaciones, es decir, de las proporciones p_1 y p_2 de individuos de dos poblaciones que presentan una determinada característica (etiquetada como *éxito*).

Para ello, se toman dos muestras aleatorias simples (una de cada población) de tamaños n_1 y n_2 y se cuenta el número de éxitos en cada una, x e y.

Estimación puntual: $\hat{p}_1 - \hat{p}_2 = \frac{x}{n_1} - \frac{y}{n_2}$

Intervalo de confianza:

$$\boxed{I_{p_1-p_2}^{1-\alpha} = \left((\hat{p}_1 - \hat{p}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} , (\hat{p}_1 - \hat{p}_2) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right)}$$

Ejemplo

Un taller utiliza dos modelos diferentes de máquina en la fabricación de una determinada pieza para un proceso productivo. Esa pieza debe tener unas medidas que cumplan ciertas especificaciones para resultar válidas. Se quiere determinar la diferencia de proporciones de piezas fabricadas por ambos modelos de máquina que no cumplan con las especificaciones y, por tanto, no sean aceptables para el proceso productivo. Para ello, se recoge una muestra aleatoria simple formada por 285 piezas fabricadas por el primer modelo de las que trece no resultan válidas. De las fabricadas por el segundo modelo de máquina se obtiene una muestra de 233 piezas de las que resultan once defectuosas. Se pide obtener un intervalo al 90% de confianza para la diferencia de proporciones de piezas que no cumplan las especificaciones de medida.

- nivel de significación y tamaños muestrales

```
α = 0.10; n1 = 285; n2 = 233;
```

- número de piezas no válidas (éxitos) en cada muestra

```
exitos1 = 13; exitos2 = 11;
```

- proporción de éxitos en cada muestra

```
p̂1 = exitos1 / n1; p̂2 = exitos2 / n2;
```

- simulación el estimador

```
zα/2 = Quantile[NormalDistribution[0, 1], 1 - α / 2];

raiz = Sqrt[p̂1 (1 - p̂1) / n1 + p̂2 (1 - p̂2) / n2];

{(p̂1 - p̂2) - zα/2 * raiz, (p̂1 - p̂2) + zα/2 * raiz}

{-0.0321836, 0.028991}
```

- solución: $\boxed{I_{p_1-p_2}^{0.90} = (-0.0321836, 0.028991)}$

- **Nota**. La proporción de piezas no válidas fabricadas por cada modelo de máquina puede considerarse igual con un 90% de confianza ya que el valor cero pertenece al intervalo calculado.