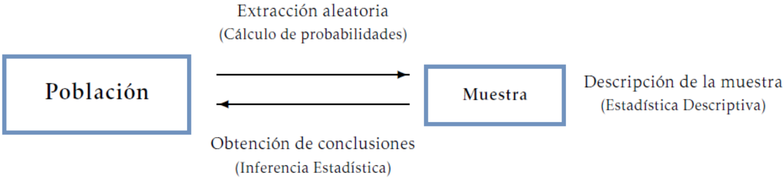
Estadística: estadística
Grado en Relaciones Laborales | Curso 2019-2020

Tema 7.A. Introdución a la inferencia estadística

Alejandro Saavedra Nieves

- Motivación.
- Principales distribuciones continuas.
- Estimación puntual.
- Contrastes de hipótesis.

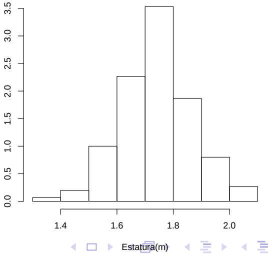
Variable: Estatura de los estudiantes de RRLL de la UVigo



Estudiantes Facultad RRLL

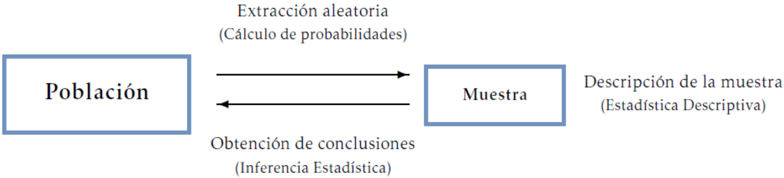
Muestra estudiantes RRLL, $n = 150$

$x_1 = 1.66$, $x_2 = 1.86$, $x_3 = 1.62$, ...



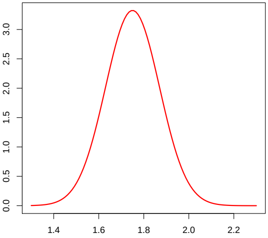
Motivación

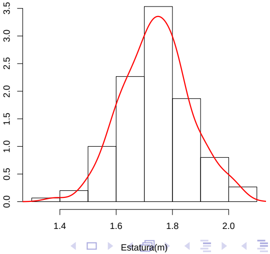
Variable: Estatura de los estudiantes de RRLL de la UVigo



Estudiantes Facultad RRLL
Modelo | Distribución normal

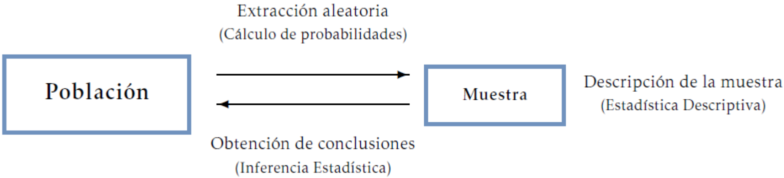
Muestra estudiantes RRLL, $n = 150$
$x_1 = 1.66$, $x_2 = 1.86$, $x_3 = 1.62$, ...



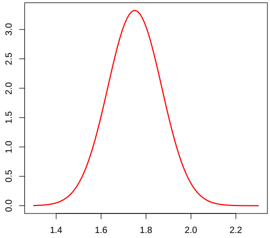


Motivación

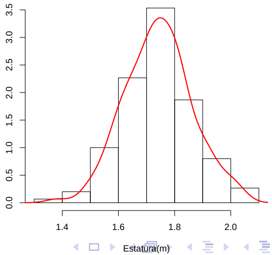
Variable: Estatura de los estudiantes de RRLL de la UVigo



Estudiantes Facultad RRLL
Modelo | Distribución normal
Media: $\mu = 1.75$; varianza: $\sigma^2 = 0.12^2$



Muestra estudiantes RRLL, $n = 150$
$x_1 = 1.66$, $x_2 = 1.86$, $x_3 = 1.62$, ...
$\overline{X} = 1.74$, $s^2 = 0.119^2$



Conceptos básicos

- **Población.** Conjunto de elementos homogéneos sobre el que queremos estudiar. Por ejemplo, el conjunto de votantes de unas elecciones.
- **Muestra.** Subconjunto representativo de la población (datos).
- **Tamaño de la muestra, n.** Número de elementos que componen la muestra.
- **Metodología de muestreo.** Procedimiento usado para la obtención de la muestra.
    - Aunque hay diferentes procedimientos para este fin, nos centraremos en el caso de **muestreo aleatorio simple** (m.a.s.).
    - De toda la población se extrae un total de n elementos. Todos los individuos de la población tienen asociada la misma probabilidad de ser extraídos.

- **Estimación Puntual**. Consiste en aventurar un valor, calculado a partir de la muestra, que esté lo más próximo posible al verdadero parámetro.
- **Intervalos de Confianza**. Dado que la estimación puntual conlleva un cierto error, construímos un intervalo que con alta probabilidad contenga al parámetro. La amplitud del intervalo nos da idea del margen de error de nuestra estimación.
- **Contrastes de Hipótesis**. Se trata de responder a preguntas muy concretas sobre la población, y se reducen a un problema de decisión sobre la veracidad de ciertas hipótesis.

- **Estimación Puntual**. Consiste en aventurar un valor, calculado a partir de la muestra, que esté lo más próximo posible al verdadero parámetro.

Introducción: conceptos básicos

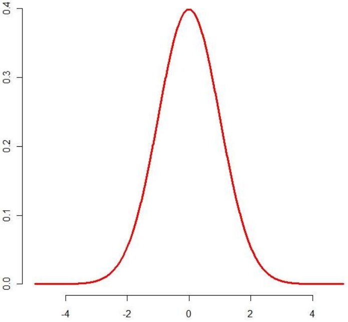
- Parámetro. Se trata de una característica medible (numérica) de la población. Por ejemplo, el número de representantes que obtiene un partido político en las elecciones.
- Espacio paramétrico. El espacio de todos los posibles valores del parámetro. En el ejemplo de las elecciones, el espacio paramétrico serán los números entre 0 y 350 (número de diputados).
- Estadístico. Una función de la muestra que no contiene valores desconocidos.
- Pivote. Una función de la muestra y de los parámetros no observados cuya distribución probabilidad es conocida ($N(0,1)$, $P(3)$, ...).

Principales modelos de distribuciones

Variable Normal Estándar

Una variable aleatoria continua Z se dice se dice que tiene distribución normal estándar, y lo denotamos $Z \in N(0,1)$, si su función de densidad viene dada por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{si} \quad z \in \mathbb{R}$$



Variable Normal Estándar

Una variable aleatoria continua Z se dice se dice que tiene distribución normal estándar, y lo denotamos $Z \in N(0,1)$, si su función de densidad viene dada por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{si} \quad z \in \mathbb{R}$$

- $Z \in N(0,1)$ toma valores en toda la recta real. $(f(z) > 0 \quad \forall z \in \mathbb{R})$
- f es simétrica en torno a cero.
- Si $Z \in N(0,1)$ entonces $\mu = 0$ y $\sigma^2 = 1$.

Distribuciones asociadas con la normal

Además del modelo normal, existen otros modelos que desempeñan un papel importante en la inferencia estadística. Entre ellos se encuentran

- la distribución χ^2
- la distribución t de Student.

Estimación puntual

Estimación puntual: introducción

El objetivo de la estimación puntual pasa por aproximar el valor de un parámetro desconocido de la población a través de un conjunto de datos de la muestra.

Ejemplo. Estimar el número de diputados que un partido conseguirá en unas elecciones.

Sea $\{X_1, \ldots, X_n\}$ una muestra obtenida mediante muestreo aleatorio simple de una variable aleatoria X.

Suponemos que la distribución de X depende de un cierto parámetro θ que puede tomar valores en algún espacio paramétrico Θ.

Ejemplo. Número de pacientes que llegan a urgencias en una hora. Esto es, $\theta = \lambda$ en una distribución $P(\lambda)$, con $\lambda \in \Theta = [0, \infty)$.

Estimación puntual: introducción

En la estimación de un mismo parámetro, se pueden usar diferentes estimadores y cada uno de ellos proporcionará una estimación diferente del parámetro en cuestión para cada muestra.

Entonces, ¿cuál es el mejor estimador?

Su respuesta no es sencilla. La idea es que nuestro estimador aproxime de manera correcta el valor real del parámetro.

Las dos propiedades de los estimadores a considerar son:

- Insesgadez.
- Consistencia.

Estimación puntual: insesgadez

Se define el sesgo de un estimador $\widehat{\theta}$ como la diferencia entre el valor esperado del estimador y el verdadero parámetro.

$$Sesgo(\widehat{\theta}) = E(\widehat{\theta}) - \theta.$$

Un estimador $\widehat{\theta}$ se dice insesgado cuando su sesgo es 0. Entonces, $E(\widehat{\theta}) = \theta$.

Esto significa que, en media, el estimador aproxima el valor real del parámetro, o equivalente, ni se infraestima ni se sobreestima el valor real del parámetro.

Estimación puntual: insesgadez

Decimos que un estimador $\widehat{\theta}$ es asintóticamente insesgado si

$$\lim_{n\to\infty} E(\widehat{\theta}) = \theta \;\Leftrightarrow\; \lim_{n\to\infty} Sesgo(\widehat{\theta}) = 0$$

Obviamente, si un estimador es insesgado, es asintóticamente insesgado.

Mientras un estimador asintóticamente insesgado puede tener sesgo, como el tamaño de la muestra aumenta y por tanto la información sobre la población, el sesgo tiende a disminuir.

Estimación puntual: insesgadez

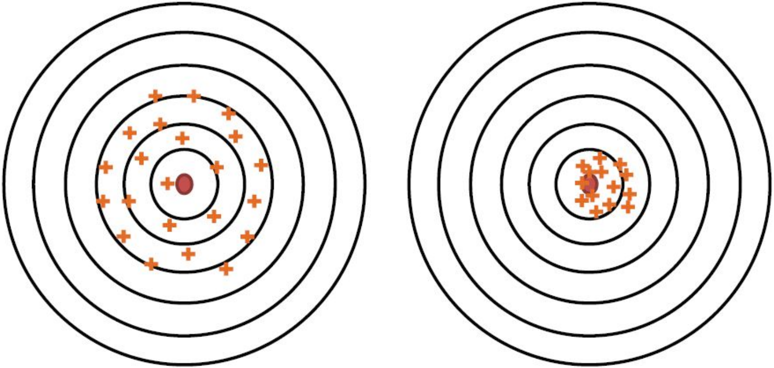
Ejemplos.

1. Los gobiernos no suelen usar estimadores insesgados cuando hacen predicciones económicas a su favor.

2. Los entrenadores de fútbol suelen no usar estimadores insesgados para referirse a la actuación arbitral. Suelen hablar de arbitraje injusto en su contra.

3. Los estudiantes aseguran que los estimadores de sus conocimientos que utilizan los profesores para calificarlos no son insesgados.

De la insesgadez a la consistencia



A la vista del ejemplo, parece que la varianza también juega un papel fundamental en la selección del mejor estimador.

Error cuadrático medio

Este error mide la distancia media entre el estimador y el valor real del parámetro.

Sea $\widehat{\theta}$ un estimador de θ. Se define el Error Cuadrático Medio (ECM) de $\widehat{\theta}$ como

$$ECM(\widehat{\theta}) = E[(\widehat{\theta} - \theta)^2].$$

Sin embargo, la expresión más común para calcular el ECM es

$$\begin{aligned} ECM(\widehat{\theta}) &= \left(E\left(\widehat{\theta}\right) - \theta\right)^2 + Var\left(\widehat{\theta}\right) \\ &= Sesgo(\widehat{\theta})^2 + Var\left(\widehat{\theta}\right) \end{aligned}$$

Nótese que en la expresión del ECM influyen tanto el sesgo y la varianza. El objetivo es buscar aquel estimador que minimice el ECM.

Consistencia

Una condición mínima de un estimador debería ser la mejora en la aproximación del parámetro cuando el tamaño muestral aumenta (más información).

Si tuviésemos infinitos datos, conoceríamos la distribución de la población y por tanto el parámetro bajo estudio.

Decimos que un estimador $\widehat{\theta}$ es consistente si las siguientes condiciones se cumplen:

- Es asintóticamente insesgado: $\lim_{n\to\infty} E\left(\widehat{\theta}\right) = \theta$
- Su varianza converge a 0 cuando n tiende infinito: $\lim_{n\to\infty} Var\left(\widehat{\theta}\right) = 0.$

Equivalentemente, si

$$\lim_{n\to\infty} ECM(\widehat{\theta}) = \lim_{n\to\infty} Sesgo\left(\widehat{\theta}\right) + \lim_{n\to\infty} Var\left(\widehat{\theta}\right) = 0.$$

$\widehat{\theta}$ es un estimador consistente.

Ejemplo

Sea X la v.a. que "tiempo en minutos entre dos llamadas consecutivas de una centralita". X sigue una distribución exponencial con parámetro $\frac{1}{\theta}$.

Dada una muestra aleatoria simple $\{X_1, \ldots, X_n\}$ con $n \geq 2$ definimos los siguientes estimadores para θ:

$$\widehat{\theta}_1 = \overline{X} \text{ y } \widehat{\theta}_2 = \frac{X_1 + X_n}{2}.$$

Nota. Si $X \sim Exp(\lambda)$, $E(X) = \frac{1}{\lambda}$ y $Var(X) = \frac{1}{\lambda^2}$

(a) Prueba que los estimadores son insesgados.

Ejemplo

Sea X la v.a. que "tiempo en minutos entre dos llamadas consecutivas de una centralita". X sigue una distribución exponencial con parámetro $\frac{1}{\theta}$.

Dada una muestra aleatoria simple $\{X_1, \dots, X_n\}$ con $n \geq 2$ definimos los siguientes estimadores para θ:

$$\widehat{\theta}_1 = \overline{X} \text{ y } \widehat{\theta}_2 = \frac{X_1 + X_n}{2}.$$

Nota. Si $X \sim Exp(\lambda)$, $E(X) = \frac{1}{\lambda}$ y $Var(X) = \frac{1}{\lambda^2}$

(a) Prueba que los estimadores son insesgados.

- $Sesgo(\widehat{\theta}_1) = E(\widehat{\theta}_1) - \theta = E(\overline{X}) - \theta = E\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) - \theta =$
  $\frac{1}{n} \cdot E\left(\sum_{i=1}^{n} X_i\right) - \theta) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i) - \theta = \frac{1}{n} \cdot n \cdot E(X) - \theta = \theta - \theta = 0.$

Ejemplo

Sea X la v.a. que "tiempo en minutos entre dos llamadas consecutivas de una centralita". X sigue una distribución exponencial con parámetro $\frac{1}{\theta}$.

Dada una muestra aleatoria simple $\{X_1, \ldots, X_n\}$ con $n \geq 2$ definimos los siguientes estimadores para θ:

$$\widehat{\theta}_1 = \overline{X} \text{ y } \widehat{\theta}_2 = \frac{X_1 + X_n}{2}.$$

Nota. Si $X \sim Exp(\lambda)$, $E(X) = \frac{1}{\lambda}$ y $Var(X) = \frac{1}{\lambda^2}$

(a) Prueba que los estimadores son insesgados.

- $Sesgo(\widehat{\theta}_1) = E(\widehat{\theta}_1) - \theta = E(\overline{X}) - \theta = E\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) - \theta =$
  $\frac{1}{n} \cdot E\left(\sum_{i=1}^{n} X_i\right) - \theta) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i) - \theta = \frac{1}{n} \cdot n \cdot E(X) - \theta = \theta - \theta = 0.$

- $Sesgo(\widehat{\theta}_2) = E(\widehat{\theta}_2) - \theta = E\left(\frac{X_1 + X_n}{2}\right) - \theta = \frac{1}{2} \cdot E(X_1 + X_2) - \theta =$
  $\frac{1}{2} \cdot 2 \cdot E(X) - \theta = \theta - \theta = 0.$

Ejemplo (continuación)

(b) ¿Cuál es el mejor?

Ejemplo (continuación)

(b) ¿Cuál es el mejor?

Dado que $ECM(\widehat{\theta}) = Sesgo(\widehat{\theta})^2 + Var(\widehat{\theta})$, tenemos que calcular las varianzas.

- $Var(\widehat{\theta}_1) = Var(\overline{X}) = Var\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n \cdot Var(X) = \frac{\theta^2}{n}$.

  Entonces, $ECM(\widehat{\theta}_1) = Sesgo(\widehat{\theta}_1)^2 + Var(\widehat{\theta}_1) = \frac{\theta^2}{n}$.

Ejemplo (continuación)

(b) ¿Cuál es el mejor?

Dado que $ECM(\widehat{\theta}) = Sesgo(\widehat{\theta})^2 + Var(\widehat{\theta})$, tenemos que calcular las varianzas.

- $Var(\widehat{\theta}_1) = Var(\overline{X}) = Var\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n \cdot Var(X) = \frac{\theta^2}{n}.$

  Entonces, $ECM(\widehat{\theta}_1) = Sesgo(\widehat{\theta}_1)^2 + Var(\widehat{\theta}_1) = \frac{\theta^2}{n}.$

- $Var\left(\widehat{\theta}_2\right) = Var\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2^2} \cdot Var(X_1 + X_2) = \frac{1}{4} \cdot [Var(X_1) + Var(X_2)] = \frac{1}{4} \cdot 2 \cdot Var(X) = \frac{\theta^2}{2}.$

  Thus, $ECM(\widehat{\theta}_2) = \frac{\theta^2}{2} \geq \frac{\theta^2}{n} = ECM(\widehat{\theta}_1)$

Ejemplo (continuación)

(b) ¿Cuál es el mejor?

Dado que $ECM(\widehat{\theta}) = Sesgo(\widehat{\theta})^2 + Var(\widehat{\theta})$, tenemos que calcular las varianzas.

- $Var(\widehat{\theta}_1) = Var(\overline{X}) = Var\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n \cdot Var(X) = \frac{\theta^2}{n}.$

  Entonces, $ECM(\widehat{\theta}_1) = Sesgo(\widehat{\theta}_1)^2 + Var(\widehat{\theta}_1) = \frac{\theta^2}{n}.$

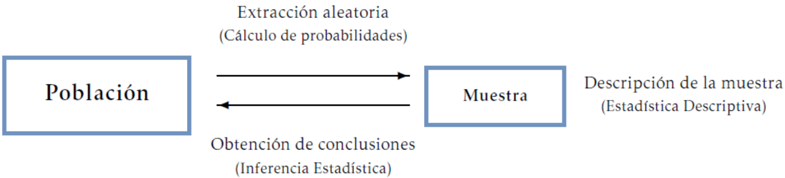
- $Var\left(\widehat{\theta}_2\right) = Var\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2^2} \cdot Var(X_1 + X_2) = \frac{1}{4} \cdot [Var(X_1) + Var(X_2)] = \frac{1}{4} \cdot 2 \cdot Var(X) = \frac{\theta^2}{2}.$

  Thus, $ECM(\widehat{\theta}_2) = \frac{\theta^2}{2} \geq \frac{\theta^2}{n} = ECM(\widehat{\theta}_1)$

Conclusion. $\widehat{\theta}_1$ es mejor estimador de θ que $\widehat{\theta}_2$.

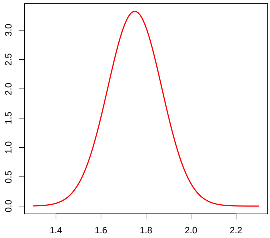
Estimación puntual (de una media)

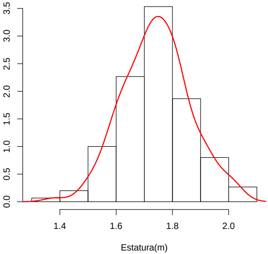
Variable: Estatura de los estudiantes de RRLL de la UVigo



Estudiantes Facultad RRLL
Modelo | Distribución normal
Media: $\mu = 1.75$; varianza: $\sigma^2 = 0.12^2$

Muestra estudiantes RRLL, $n = 150$
$x_1 = 1.66$, $x_2 = 1.86$, $x_3 = 1.62$, ...
$\overline{X} = 1.74$, $s^2 = 0.119^2$





Estimación puntual (de una media)

Objetivo. Estimar la media de una característica en una población.

Sea $X_1, X_2, \ldots, X_n$ una m.a.s. con $X_i \sim N(\mu, \sigma)$.

Un estimador puntual para la media $\boldsymbol{\mu}$ es:

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

- Insesgado. $E(\overline{X}) = E(X) = \mu$.
- Consistente. $Var(\overline{X}) = \dfrac{Var(X)}{n} = \dfrac{\sigma^2}{n}$.

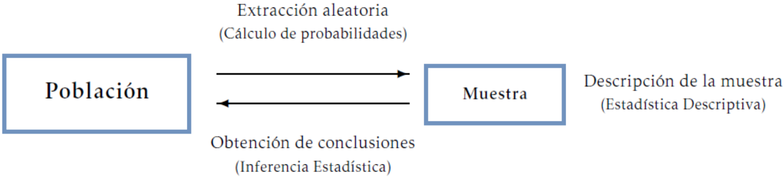
Entonces,

Distribución de $\bar{X}$

$$\bar{X} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

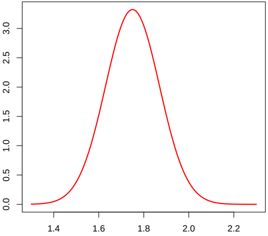
Estimación puntual (de una varianza)

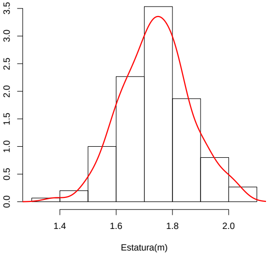
Variable: Estatura de los estudiantes de RRLL de la UVigo



Estudiantes Facultad RRLL
Modelo | Distribución normal
Varianza: $\sigma^2 = 0.12^2$

Muestra estudiantes RRLL, $n = 150$
$x_1 = 1.66$, $x_2 = 1.86$, $x_3 = 1.62$, ...
$S_X^2 = 0.119^2$





Estimación puntual (de una varianza) con μ conocida

Objetivo. Estimar la varianza de una característica en una población con μ conocida.

Sea $X_1, X_2, \ldots, X_n$ una m.a.s. extraida de una variable aleatoria X, con $E(X) = \mu$ y $Var(X) = \sigma^2$.

Un estimador puntual para la varianza σ^2 es:

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

- Asintóticamente insesgado. $E(S^2) = \dfrac{n-1}{n}\sigma^2$.
- Consistente.

Entonces,

Distribución de S^2

$$\frac{nS^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n.$$

Estimación puntual (de una varianza) con μ desconocida

Objetivo. Estimar la varianza de una característica en una población con μ desconocida.

Sea $X_1, X_2, \ldots, X_n$ una m.a.s. extraida de una variable aleatoria X, con $E(X) = \mu$ (desconocida) y $Var(X) = \sigma^2$.

Un estimador puntual para la varianza σ^2 es:

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

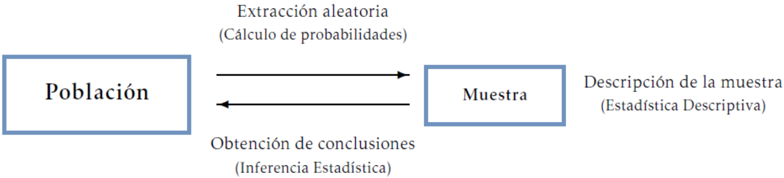
- Insesgado. $E(S_c^2) = \sigma^2$.
- Consistente.

Entonces,

Distribución de S_c^2

$$\frac{(n-1)S_c^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Variable: Estatura de los estudiantes de RRLL de la UVigo



Estudiantes Facultad RRLL

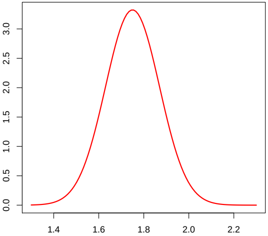
Modelo | Distribución normal

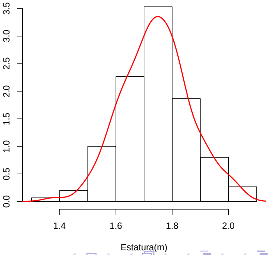
$p = P(X > 1.75)$

Muestra estudiantes RRLL, $n = 150$

$x_1 = 1.66,\ x_2 = 1.86,\ x_3 = 1.62,\ \ldots$

$\hat{p} = \frac{\#\{x_i > 1.75\}}{n}$





Estimación puntual (de una proporción)

Objetivo. Estimar la proporción de elementos p en una población que presentan alguna característica.

Dada una m.a.s., $X_i = 1$ es igual a 1 si el elemento presenta la característica y 0, en otro caso. Entonces, cada variable en la muestra es $X_i \sim Be(p)$.

Un estimador para p será la proporción muestral:

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

- Insesgado. $E(\widehat{p}) = p$.
- Consistente. $Var(\widehat{p}) = Var(\overline{X}) = \frac{p(1-p)}{n}$.

Para $n > 0$, por el Teorema Central de Límite:

Distribución de $\hat{p}$

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Intervalos de confianza

- **Intervalos de Confianza**. Dado que la estimación puntual conlleva un cierto error, construímos un intervalo que con alta probabilidad contenga al parámetro. La amplitud del intervalo nos da idea del margen de error de nuestra estimación.

- Un intervalo de confianza es un intervalo construido en base a la muestra y, por tanto, aleatorio, que contiene al parámetro con una cierta probabilidad, conocida como **nivel de confianza**.

Intervalo de confianza

- Sea θ el parámetro desconocido y $\alpha \in [0, 1]$.
- Se dice que el intervalo $[L_1, L_2]$ tiene un nivel de confianza $1 - \alpha$ si

$$P(L_1 \leq \theta \leq L_2) \geq 1 - \alpha$$

Intervalo de confianza

- Sea θ el parámetro desconocido y $\alpha \in [0, 1]$.
- Se dice que el intervalo $[L_1, L_2]$ tiene un nivel de confianza $1 - \alpha$ si

$$P(L_1 \leq \theta \leq L_2) \geq 1 - \alpha$$

- Los valores de L_1 y L_2 **dependerán de la muestra!!!!**.

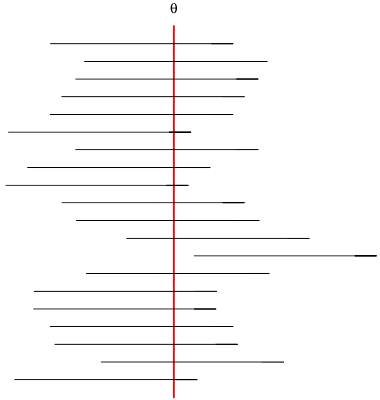
Intervalo de confianza

- Sea θ el parámetro desconocido y $\alpha \in [0, 1]$.
- Se dice que el intervalo $[L_1, L_2]$ tiene un nivel de confianza $1 - \alpha$ si

$$P(L_1 \leq \theta \leq L_2) \geq 1 - \alpha$$

- Los valores de L_1 y L_2 **dependerán de la muestra!!!!**.
- El nivel de confianza con frecuencia se expresa en porcentaje. Así, un intervalo de confianza del 95% es un intervalo de extremos aleatorios que contiene al parámetro con una probabilidad de 0.95.

Interpretación del nivel de confianza $1-\alpha$



- Dada una realización muestral, el intervalo construido puede contener o no al parámetro desconocido
- Esperamos que el $100(1-\alpha)\%$ de los intervalos contengan al parámetro desconocido

Intervalo de confianza para la media μ de una población normal (σ^2 conocida)

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$

- Recordamos que es este caso

$$\boxed{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \in N(0,1)}$$

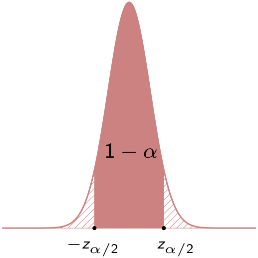
- Este estadístico (pivote) nos servirá para construir un I.C. con nivel de confianza $1 - \alpha$ para la media μ cuando la **varianza σ^2 es conocida**.

Intervalo de confianza para la media μ de una población normal (σ^2 conocida)

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$. Supongamos que σ^2 es conocida.

- Sea $z_{\alpha/2}$ el valor tal que $P(Z > z_{\alpha/2}) = \alpha/2$, siendo $Z \in N(0,1)$. Entonces:

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



Intervalo de confianza para la media μ de una población normal (σ^2 conocida)

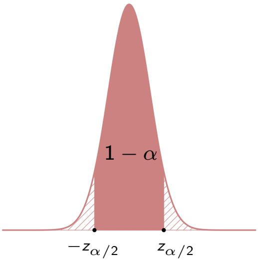
Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$. Supongamos que σ^2 es conocida.

- Sea $z_{\alpha/2}$ el valor tal que $P(Z > z_{\alpha/2}) = \alpha/2$, siendo $Z \in N(0,1)$. Entonces:

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

- Equivalentemente,

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



Intervalo de confianza para la media μ de una población normal (σ^2 conocida)

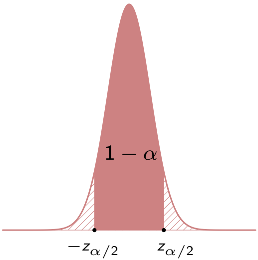
Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$. Supongamos que σ^2 es conocida.

- Sea $z_{\alpha/2}$ el valor tal que $P(Z > z_{\alpha/2}) = \alpha/2$, siendo $Z \in N(0,1)$. Entonces:

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

- Equivalentemente,

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



I.C. de nivel $1-\alpha$ para la media μ cuando σ^2 es conocida

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

Intervalo de confianza para la media μ de una población normal (σ^2 conocida)

I.C. de nivel $1 - \alpha$ para la media μ cuando σ^2 es conocida

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

Ejercicio: Estamos interesados en determinar la estatura media (m) de los alumnos de la Facultad de RRLL. Para ello toma una muestra de 10 alumnos y determina la estatura de cada uno de ellos de cada uno de ellos. Los resultados son los siguientes:

1.66 1.86 1.62 1.74 1.54 1.78 1.90 1.80 1.58 1.60

- ¿Cómo estimarías la altura media a partir de esta muestra?
- Se sabe que la variable de interés es aproximadamente normal con varianza igual a 0.12^2. Construye un intervalo de confianza para la estatura media con nivel de confianza del 95%.
- ¿Cuál sería el intervalo de confianza para un nivel de confianza del 90%?

Intervalo de confianza para la media μ de una población normal (σ^2 desconocida)

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$.

- En la práctica no es habitual conocer la varianza de la variable de interés.

Intervalo de confianza para la media μ de una población normal (σ^2 desconocida)

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$.

- En la práctica no es habitual conocer la varianza de la variable de interés.
- Cuando la varianza σ^2 es desconocida, usaremos como estadístico (pivote) para construir un intervalo de confianza para la media μ

$$\frac{\bar{X} - \mu}{S_c/\sqrt{n}}$$

- Recuerda que:

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

Intervalo de confianza para la media μ de una población normal (σ^2 desconocida)

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$.

- En la práctica no es habitual conocer la varianza de la variable de interés.
- Cuando la varianza σ^2 es desconocida, usaremos como estadístico (pivote) para construir un intervalo de confianza para la media μ

$$\frac{\bar{X} - \mu}{S_c/\sqrt{n}}$$

- Recuerda que:

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

- En este caso:

$$\boxed{\frac{\bar{X} - \mu}{S_c/\sqrt{n}} \in t_{n-1}}$$

Intervalo de confianza para la media μ de una población normal (σ^2 desconocida)

I.C. de nivel $1-\alpha$ para la media μ cuando σ^2 es desconocida

$$\left(\bar{X} - t_{n-1,\alpha/2}\,\frac{S_c}{\sqrt{n}}\;,\;\bar{X} + t_{n-1,\alpha/2}\,\frac{S_c}{\sqrt{n}}\right)$$

Ejercicio: Estamos interesados en determinar la estatura media (m) de los alumnos de la Facultad de RRLL. Para ello toma una muestra de 10 alumnos y determina la estatura de cada uno de ellos de cada uno de ellos. Los resultados son los siguientes:

1.66 1.86 1.62 1.74 1.54 1.78 1.90 1.80 1.58 1.60

- ¿Cómo estimarías la altura media a partir de esta muestra?
- Supongamos que la variable de interés tiene una varianza desconocida. Construye un intervalo de confianza para la estatura media con nivel de confianza del 95%.
- ¿Cuál sería el intervalo de confianza para un nivel de confianza del 90%?

Intervalo de confianza para la varianza σ^2 de una población normal con μ conocida

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$.

- Para establecer un I.C. para la varianza, usaremos como estadístico (pivote) para construir un intervalo de confianza para σ, con μ conocida:

$$\frac{nS^2}{\sigma^2}$$

- Recuerda que:

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

Intervalo de confianza para la varianza σ^2 de una población normal con μ conocida

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$.

- Para establecer un I.C. para la varianza, usaremos como estadístico (pivote) para construir un intervalo de confianza para σ, con μ conocida:

$$\frac{nS^2}{\sigma^2}$$

- Recuerda que:

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

- En este caso:

$$\boxed{\frac{nS^2}{\sigma^2} \sim \chi^2_n}$$

Intervalo de confianza para la varianza σ^2 de una población normal con μ conocida

I.C. de nivel $1 - \alpha$ para la varianza σ^2 con μ conocida

$$\left(\frac{nS^2}{\chi^2_{n,\alpha/2}}, \frac{nS^2}{\chi^2_{n,1-\alpha/2}} \right)$$

Ejercicio: Con el fin de garantizar la salubridad de cierto alimento se ha llevado a cabo un análisis que cuenta el número de bacterias que se encuentran en 9 unidades del mismo.

Los recuentos han dado los siguientes resultados:

$$157, 186, 179, 163, 171, 154, 177, 165, 168.$$

Aceptando normalidad en la distribución del recuento bacteriano, calcula el intervalo de confianza para la varianza a un nivel del 95%, sabiendo que $\mu = 170$.

Intervalo de confianza para la varianza σ^2 de una población normal con μ desconocida

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$.

- Para establecer un I.C. para la varianza, usaremos como estadístico (pivote) para construir un intervalo de confianza para σ, con μ desconocida:

$$\frac{(n-1)S_c^2}{\sigma^2}$$

- Recuerda que:

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

Intervalo de confianza para la varianza σ^2 de una población normal con μ desconocida

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $N(\mu, \sigma)$.

- Para establecer un I.C. para la varianza, usaremos como estadístico (pivote) para construir un intervalo de confianza para σ, con μ desconocida:

$$\frac{(n-1)S_c^2}{\sigma^2}$$

- Recuerda que:

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

- En este caso:

$$\boxed{\frac{(n-1)S_c^2}{\sigma^2} \sim \chi^2_{n-1}}$$

Intervalo de confianza para la varianza σ^2 de una población normal con μ desconocida

I.C. de nivel $1-\alpha$ para la varianza σ^2 con μ desconocida

$$\left(\frac{(n-1)S_c^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_c^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

Ejercicio: Con el fin de garantizar la salubridad de cierto alimento se ha llevado a cabo un análisis que cuenta el número de bacterias que se encuentran en 9 unidades del mismo.
Los recuentos han dado los siguientes resultados:

$$157, 186, 179, 163, 171, 154, 177, 165, 168.$$

Aceptando normalidad en la distribución del recuento bacteriano, calcula el intervalo de confianza para la media y varianza a un nivel del 95%.

Intervalo de confianza para una proporción

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $X_i \sim Be(p)$.

- En este caso, usaremos como estadístico (pivote) para construir un intervalo de confianza para la proporción p
$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$
- Recuerda que:
$$\widehat{p} = \frac{X_1 + \cdots + X_n}{n}$$

Intervalo de confianza para una proporción

Sea $X_1, X_2, \ldots, X_n$ una muestra formada por n variables independientes y con la misma distribución $X_i \sim Be(p)$.

- En este caso, usaremos como estadístico (pivote) para construir un intervalo de confianza para la proporción p
$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$
- Recuerda que:
$$\widehat{p} = \frac{X_1 + \cdots + X_n}{n}$$
- En este caso:
$$\boxed{\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \in N(0,1)}$$

Intervalo de confianza para una proporción

I.C. de nivel $1 - \alpha$ para una proporción

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

Ejercicio: En una encuesta realizada a 150 familias de una determinada población, se encontró que en 25 de ellas había tres o más hijos.

Halla el intervalo de confianza para estimar la proporción real de las familias en las que hay tres o más hijos, con un nivel de confianza del 90%.

Cálculo del tamaño de muestra necesario

Nota: Este razonamiento será aplicable todos los procedimientos mostrados hasta el momento.

Sólo lo razonaremos para el caso del I.C. de la media.

- Hasta este momento, dada una muestra de tamaño n, el I.C. para la media nos determina la precisión en nuestra estimación en términos de la longitud del intervalo de confianza.

- Sin embargo, fijada una longitud del intervalo deseada, podemos determinar el tamaño muestral n necesario para este fin.

- Este se obtiene despejando n en la expresión de la longitud del intervalo (L)

$$L = L_2 - L_1 = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Esto es, $n = \frac{4z_{\alpha/2}^2\sigma^2}{L^2}$.

Intervalo de confianza para la diferencia de medias (σ^2 conocida)

Objetivo. En ocasiones, queremos comparar si dos poblaciones diferentes tienen la misma media determinada característica.

Sea $X_1, X_2, \ldots, X_{n_1}$ una muestra de tamaño n_1 de una normal $N(\mu_1, \sigma_1)$ e $Y_1, Y_2, \ldots, Y_{n_2}$ una muestra de tamaño n_2 de una $N(\mu_2, \sigma_2)$

- Cuando la varianza σ_1^2 y σ_2^2 son conocidas, usaremos como estadístico (pivote) para construir un intervalo de confianza para la diferencia de medias $\mu_1 - \mu_2$

$$\frac{(\bar{X} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- En este caso:

$$\boxed{\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \in N(0,1)}$$

I.C. de nivel $1 - \alpha$ para la diferencia de medias

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\ ,\ \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Diremos que las medias son iguales para ese α si el valor 0 está en el intervalo.

Ejercicio: 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 $\sigma_1^2 = 20^2$ y $\sigma_2^2 = 25^2$, $\bar{x} = 82.5$, $\bar{y} = 85.4$, $n_1 = 70$ y $n_2 = 50$.

Calcula el correspondiente I.C. para la diferencia de medias con una confianza del 95%.

Intervalo de confianza para la diferencia de medias (σ^2 desconocida)

Sea $X_1, X_2, \ldots, X_{n_1}$ una muestra de tamaño n_1 de una normal $N(\mu_1, \sigma_1)$ e $Y_1, Y_2, \ldots, Y_{n_2}$ una muestra de tamaño n_2 de una $N(\mu_2, \sigma_2)$

- Cuando la varianza σ_1^2 y σ_2^2 son desconocidas, usaremos como estadístico (pivote) para construir un intervalo de confianza para la diferencia de medias $\mu_1 - \mu_2$

$$\frac{(\bar{X} - \bar{y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

con

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

- En este caso:

$$\boxed{\frac{(\bar{X} - \bar{y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \in t_{n_1+n_2-2}}$$

Intervalo de confianza para la diferencia de medias (σ^2 desconocida)

I.C. de nivel $1-\alpha$ para la diferencia de medias (σ^2 desconocida)

$$\left(\bar{X}-\bar{Y}-t_{n_1+n_2-2,\alpha/2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\ ,\ \bar{X}-\bar{Y}+t_{n_1+n_2-2,\alpha/2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\ \right)$$

Diremos que las medias son iguales para ese α si el valor 0 está en el intervalo.

Ejercicio: 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 son desconocidas, pero que $S_1^2 = 21.4^2$, $S_2^2 = 24.3^2$, $\bar{x} = 82.5$, $\bar{y} = 85.4$, $n_1 = 70$ y $n_2 = 50$.

Calcula el correspondiente I.C. para la diferencia de medias con una confianza del 95%.

Intervalo de confianza para el cociente de varianzas

Sea $X_1, X_2, \ldots, X_{n_1}$ una muestra de tamaño n_1 de una normal $N(\mu_1, \sigma_1)$ e $Y_1, Y_2, \ldots, Y_{n_2}$ una muestra de tamaño n_2 de una $N(\mu_2, \sigma_2)$

- En este caso, usaremos como estadístico (pivote) para construir un intervalo de confianza para el cociente de varianzas $\frac{\sigma_1^2}{\sigma_2^2}$.

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

- En este caso:

$$\boxed{\frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \in F_{n_1-1, n_2-1}}$$

Intervalo de confianza para el cociente de varianzas

I.C. de nivel $1 - \alpha$ para el cociente de varianzas

$$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1, n_2-1, \alpha/2}}, \frac{S_1^2}{S_2^2} F_{n_1-1, n_2-1, \alpha/2} \right)$$

Diremos que las medias son iguales para ese α si el valor 1 está en el intervalo.

Ejercicio: Siguiendo con el ejemplo de los alquileres en Madrid y en Barcelona, calcula un intervalo de confianza para el cociente de varianzas.

En el futuro (no lejano)...

- En ocasiones, plantearemos hipótesis acerca de nuestro parámetro de interés.
- Usaremos procedimientos estadísticos para rechazar o aceptar esta suposición.
- Pero la decisión que tomemos no tiene que ser cierta...