>>> Modelos Estadísticos vs Algoritmos
>>> *¿Juntos pero no revueltos?*

Name: Joaquín Ureta A.[†]
Date: July 13, 2020

[†]Co-fundador Moebius Analítica | joaquin@moebius-analitica.cl

>>> Lo que veremos hoy

1. ¿Qué es la Estadística? Perspectiva Frecuentista vs Bayesiana

2. Definición de un Modelo Estadístico

3. Inferencia Estadística aplicada a regresiones

\>>> ¿Qué es la Estadística?

La Estadística es la rama de las matemáticas que estudia la *variabilidad*, así como el proceso aleatorio que la genera siguiendo las *leyes de la probabilidad*. [...] la estadística es una *ciencia formal deductiva*, con un conocimiento propio, dinámico y en continuo desarrollo obtenido a través del método científico formal. [1]

[1] https://en.wikipedia.org/wiki/Statistics

>>> ¿Qué es la Estadística?

Para estudiar cualquier fenómeno siempre contamos con una porción de información (subconjunto de datos del universo). Por tanto, ¿cómo podemos lidiar con ello? La Estadística nos da una mano para el entendimiento del fenómeno e inferir su comportamiento futuro.

>>> Perspectiva frecuentista de la Estadística

Dentro de la Estadística existen dos corrientes: Frecuentista y Bayesiana.

[2] R. Fisher fue quién le dio una *matematización* a la Estadística frecuentista moderna.

>>> Perspectiva frecuentista de la Estadística

Dentro de la Estadística existen dos corrientes: Frecuentista y Bayesiana.

La Estadística *frecuentista*[2] razona en base a la replicabilidad de un fenómeno (experimentación), para así estimar la probabilidad de un evento (lo que llamamos frecuencia relativa). A medida que pudiésemos replicar el experimento un número infinito de veces, aquella probabilidad basada en la frecuencia relativa tendería a ser la probabilidad real del evento, i.e.,

$$\lim_{N\to\infty}\left(\frac{n(E)}{N}\right) = \mathbb{P}(E), \tag{1}$$

donde E representa un evento dentro de un experimento, N el tamaño muestral y $\mathbb{P}(\cdot)$ representa una medida de probabilidad.

[2] R. Fisher fue quién le dio una *matematización* a la Estadística frecuentista moderna.

>>> Perspectiva bayesiana de la Estadística

Ahora, ya hemos señalado que siempre tendremos una mirada *parcial* de la realidad, por tanto la posición frecuentista da por sentado, en base a supuestos asintóticos, que los resultados obtenidos son *reales*.

>>> Perspectiva bayesiana de la Estadística

Ahora, ya hemos señalado que siempre tendremos una mirada *parcial* de la realidad, por tanto la posición frecuentista da por sentado, en base a supuestos asintóticos, que los resultados obtenidos son *reales*.

La posición bayesiana, por el contrario, nos viene a cuestionar este hecho. Para un bayesiano, la asignación de probabilidad de un evento se debe al conocimiento limitado de la realidad, i.e. la incertidumbre inherente. Esto nos plantea la siguiente cuestión: a medida que yo almaceno más evidencia (nuevos datos), puedo ir refinando (actualizar) la probabilidad del evento, debido a la reducción de la incertidumbre del mismo. El principio antes descrito se basa en lo siguiente:

>>> Perspectiva bayesiana de la Estadística

Theorem (Bayes)

Sea $\{A_j\}_{j=1}^{k}$ una partición de un espacio muestral Ω, con probabilidades *a priori* $\mathbb{P}(A_j)$. Sea B cualquier evento tal que $\mathbb{P}(B) > 0$, entonces la probabilidad *a posteriori* de A_j está dada por:

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)}{\mathbb{P}(B)}, \forall j \in \{1, 2, \ldots, k\} \tag{2}$$

>>> Comparación entre ellas

Dada la exposición basal de ambas corrientes, sería natural preguntarse cuál usar... La respuesta a esa pregunta es

[3] Por ejemplo, algoritmo de Metrópolis-Hasting y MCMC

>>> Comparación entre ellas

Dada la exposición basal de ambas corrientes, sería natural preguntarse cuál usar... La respuesta a esa pregunta es AMBAS!!

[3] Por ejemplo, algoritmo de Metrópolis-Hasting y MCMC

>>> Comparación entre ellas

Dada la exposición basal de ambas corrientes, sería natural preguntarse cuál usar... La respuesta a esa pregunta es AMBAS!! Si bien la perspectiva frecuentista descansa en supuestos más complejos de alcanzar, ésta ha sido exitosa durante todo el siglo pasado (y aún).
Por otro lado, la Estadística bayesiana ha tomado una fuerza importante gracias a los avances computacionales durante el siglo recién pasado. En general, el cálculo de las probabilidades en (2) han podido aproximarse a través del desarrollo de técnicas de simulación [3]

[3] Por ejemplo, algoritmo de Metrópolis-Hasting y MCMC

>>> Definición de un Modelo Estadístico

OK, asentemos conceptos con la siguiente definición:

>>> Definición de un Modelo Estadístico

OK, asentemos conceptos con la siguiente definición:

Definition

Un Modelo Estadístico $\mathcal{E}$ se define como:

$$\mathcal{E} = \{(\Omega, \mathcal{A}), P^{\theta} : \theta \in \Theta\} \tag{3}$$

donde,

>>> Definición de un Modelo Estadístico

OK, asentemos conceptos con la siguiente definición:

Definition

Un Modelo Estadístico $\mathcal{E}$ se define como:

$$\mathcal{E} = \{(\Omega, \mathcal{A}), P^\theta : \theta \in \Theta\} \tag{3}$$

donde,

* $(\Omega, \mathcal{A})$ es el *espacio de muestreo*, con Ω siendo el *espacio muestral* y $\mathcal{A} \subset \mathcal{P}(\mathcal{A})$ es una σ-álgebra y representa el conjunto de eventos de interés;

>>> Definición de un Modelo Estadístico

OK, asentemos conceptos con la siguiente definición:

Definition

Un Modelo Estadístico $\mathcal{E}$ se define como:

$$\mathcal{E} = \{(\Omega, \mathcal{A}), P^{\theta} : \theta \in \Theta\} \tag{3}$$

donde,

* $(\Omega, \mathcal{A})$ es el *espacio de muestreo*, con Ω siendo el *espacio muestral* y $\mathcal{A} \subset \mathcal{P}(\mathcal{A})$ es una σ-álgebra y representa el conjunto de eventos de interés;
* θ representa un índice denominado *parámetro*;

>>> Definición de un Modelo Estadístico

OK, asentemos conceptos con la siguiente definición:

Definition

Un Modelo Estadístico $\mathcal{E}$ se define como:

$$\mathcal{E} = \{(\Omega, \mathcal{A}), P^{\theta} : \theta \in \Theta\} \tag{3}$$

donde,

* $(\Omega, \mathcal{A})$ es el *espacio de muestreo*, con Ω siendo el *espacio muestral* y $\mathcal{A} \subset \mathcal{P}(\mathcal{A})$ es una σ-álgebra y representa el conjunto de eventos de interés;
* θ representa un índice denominado *parámetro*;
* Θ, llamado *espacio paramétrico*, es un conjunto no vacío de parámetros;

>>> Definición de un Modelo Estadístico

OK, asentemos conceptos con la siguiente definición:

Definition

Un Modelo Estadístico $\mathcal{E}$ se define como:

$$\mathcal{E} = \{(\Omega, \mathcal{A}), P^{\theta} : \theta \in \Theta\} \tag{3}$$

donde,

* $(\Omega, \mathcal{A})$ es el *espacio de muestreo*, con Ω siendo el *espacio muestral* y $\mathcal{A} \subset \mathcal{P}(\mathcal{A})$ es una σ-álgebra y representa el conjunto de eventos de interés;
* θ representa un índice denominado *parámetro*;
* Θ, llamado *espacio paramétrico*, es un conjunto no vacío de parámetros;
* $P^{\bullet}(\bullet) : \Theta \times \mathcal{A} \to [0, 1]$, es una función tal que para cada $\theta \in \Theta$, $P^{\theta}(\bullet)$ es una probabilidad definida sobre $(\Omega, \mathcal{A})$; se llama *probabilidad de muestreo*.

>>> Definición de un Modelo Estadístico

La diferencia entre la estadística Bayesiana y la Estadística frecuentista es que la primera razona usando una *única medida de probabilidad* definida sobre el espacio producto "observaciones × parámetros" , mientras que la última lo hace con una *familia de probabilidades de muestreo*.

>>> Definición de un Modelo Estadístico

La diferencia entre la estadística Bayesiana y la Estadística frecuentista es que la primera razona usando una *única medida de probabilidad* definida sobre el espacio producto "observaciones × parámetros", mientras que la última lo hace con una *familia de probabilidades de muestreo*.

Un ejemplo de la definición anterior es el siguiente:

* Supongamos un modelo normal $X \sim N(\mu, \sigma^2)$. El experimento estadístico subyacente está dado por:

$$\mathcal{E} = \{(\mathbb{R}, \mathcal{B}), N(\mu, \sigma^2) : (\mu, \sigma^2) \in (\mathbb{R} \times \mathbb{R}_+)\} \tag{4}$$

  Aquí, el parámetro está dado por $\theta = (\mu, \sigma^2)$, el cual indexa a las probabilidades de muestreo que son distribuciones normales. El espacio paramétrico está dado por $\Theta = \mathbb{R} \times \mathbb{R}_+$ (subespacio vectorial de dimensión finita).

>>> Algunas observaciones (¡no se asusten!)

Es prudente hacer algunas observaciones antes de continuar:

* Debemos tener claridad de que nuestro modelo estadístico sea *identificado*. En términos específicos, un modelo está indetificado si:

$$f := \theta \to P^{\theta}$$

es una función inyectiva.

>>> Algunas observaciones (¡no se asusten!)

Es prudente hacer algunas observaciones antes de continuar:

* Debemos tener claridad de que nuestro modelo estadístico sea *identificado*. En términos específicos, un modelo está indetificado si:

  $$f := \theta \to P^{\theta}$$

  es una función inyectiva.
* Debemos ser capaces de resolver el problema de *especificación* de un modelo estadístico, esto es identificar la forma matemática de las distribuciones de probabilidad que generan la población hipotética de estudio.

>>> Algunas observaciones (¡no se asusten!)

Es prudente hacer algunas observaciones antes de continuar:

* Debemos tener claridad de que nuestro modelo estadístico sea *identificado*. En términos específicos, un modelo está indetificado si:

  $$f := \theta \to P^{\theta}$$

  es una función inyectiva.
* Debemos ser capaces de resolver el problema de *especificación* de un modelo estadístico, esto es identificar la forma matemática de las distribuciones de probabilidad que generan la población hipotética de estudio.
* Debemos definir qué metodo utilizaremos para realizar la estimación de los parámetros desde la muestra y que sea extrapolable a la población infinita. Éste es el problema de *estimación*.

>>> Inferencia Estadística en general

Ufff... Al parecer un modelo estadístico es más que una fórmula matemática, ¿no?

>>> Inferencia Estadística en general

Ufff... Al parecer un modelo estadístico es más que una fórmula matemática, ¿no?
Ya que hemos definido lo que es un modelo estadístico, podemos *jugar* a realizar lo esencial de la Estadística: hacer *Inferencia*.

Definition

La *Inferencia Estadística* es una parte de la Estadística que se encarga de utilizar *métodos deductivos* con el objetivo de obtener *conclusiones útiles* para hacer deducciones sobre una totalidad, basándose en la información contenida en una muestra.

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

* Paso N°1: Definición y planteamiento de un problema.

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

* Paso N°1: Definición y planteamiento de un problema.
* Paso N°2: Realizar muestreo.

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

* Paso N°1: Definición y planteamiento de un problema.
* Paso N°2: Realizar muestreo.
* Paso N°3: Definición del modelo estadístico.

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

* Paso N°1: Definición y planteamiento de un problema.
* Paso N°2: Realizar muestreo.
* Paso N°3: Definición del modelo estadístico.
* Paso N°4: Procesamiento de los datos.

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

* Paso N°1: Definición y planteamiento de un problema.
* Paso N°2: Realizar muestreo.
* Paso N°3: Definición del modelo estadístico.
* Paso N°4: Procesamiento de los datos.
* Paso N°5: Estimación de los parámetros del modelo estadístico.

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

* Paso N°1: Definición y planteamiento de un problema.
* Paso N°2: Realizar muestreo.
* Paso N°3: Definición del modelo estadístico.
* Paso N°4: Procesamiento de los datos.
* Paso N°5: Estimación de los parámetros del modelo estadístico.
* Paso N°6: Contrastación de hipótesis.

>>> Método Inferencial: paso a paso

Para hacer Inferencia, debemos realizar una serie de pasos que involucran el problema de investigación en sí, pasando por técnicas de estimación entre otros. Veamos cómo se construye:

* Paso N°1: Definición y planteamiento de un problema.
* Paso N°2: Realizar muestreo.
* Paso N°3: Definición del modelo estadístico.
* Paso N°4: Procesamiento de los datos.
* Paso N°5: Estimación de los parámetros del modelo estadístico.
* Paso N°6: Contrastación de hipótesis.
* Paso final: Concluir!

>>> Modelos de regresión

En simple, un modelo de regresión es el establecimiento de alguna relación (lineal o no lineal) entre un conjunto de variables independientes (covariables) y una variable independiente (u objetivo).
Si existiese una fuerte evidencia de alguna relación, ¿se podría inferir causalidad?

>>> Modelos de regresión

En simple, un modelo de regresión es el establecimiento de alguna relación (lineal o no lineal) entre un conjunto de variables independientes (covariables) y una variable independiente (u objetivo).
Si existiese una fuerte evidencia de alguna relación, ¿se podría inferir causalidad?
NO!!!!!!!!!!!!!!! En Estadística NUNCA podemos deducir causalidad!

>>> Modelo de regresión lineal normal

Para asentar ideas, investiguemos el modelo de regresión más utilizado en la historia: Modelo de Regresión Lineal Normal.

La especificación de este modelo es la siguiente:

$$Y = \beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_k \cdot x_k + \epsilon$$

donde $\epsilon \sim N(0, \sigma^2)$ representa un vector de errores iid.
Esto establece la siguiente relación:

$$\mathbb{E}_{\widetilde{\beta}}(Y|X, \sigma^2) = \beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_k \cdot x_k$$

donde $\widetilde{\beta}$ representa el vector de parámetros.

>>> Supuestos del modelo

Si bien la forma de este modelo supone una aparante *inocencia*, los supuestos subyacentes en él son potentes:

>>> Supuestos del modelo

Si bien la forma de este modelo supone una aparante *inocencia*, los supuestos subyacentes en él son potentes:

* Se materializa la incerteza del modelo a través de la distribución de ϵ;
* La distribución de Y se impone como una distribución normal, con parámetros *fijos* $\widetilde{\beta}$ (desconocido) y σ^2 (conocido);
* La varianza σ^2 es fija y conocida (hasta ahora).
* La distribución de $\widetilde{x}$ es fija en el momento exacto que las observo.

>>> Métodos de estimación

Ya que tenemos caracterizado nuestro modelo estadístico, ahora debemos encontrar los parámetros asociados al modelo. En esta etapa, el problema se transforma de un problema estadístico a un problema de optimización.

Definition

Estimador Máximo Verosímil. Dado un conjunto de n datos provenientes de una muestra aleatoria simple, $\{x\}_{j=1}^{n}$, el estimador máximo verosímil, $\widehat{\widetilde{\beta}}$, viene dado por:

$$\widehat{\widetilde{\beta}} = argmax \mathbb{P}(x|\widetilde{\beta})$$