

# Inferencia

Rodrigo Zepeda-Tello y Luis Carlos Bernal

2021-01-12



# Contents

<b>1</b>	<b>Introducción</b>	<b>5</b>
1.1	Estadística y muestras . . . . .	5
1.2	Modelos . . . . .	8



# Chapter 1

## Introducción

### 1.1 Estadística y muestras

La enciclopedia Stanford de filosofía establece la siguiente definición de estadística<sup>1</sup>.

**Estadística** La estadística es una disciplina matemática y conceptual que se enfoca en la relación entre datos e hipótesis. Los datos son registros de observaciones o eventos en un estudio científico, por ejemplo, un conjunto de mediciones de individuos de una población. Los datos que son obtenidos se conoce como la muestra, datos muestrales, o simplemente los datos, y todas las posibles muestras posibles en un estudio forman una colección llamada el espacio muestra. Las hipótesis, por su parte, son enunciados generales sobre el sistema objetivo de la investigación científica, por ejemplo, expresar un hecho general sobre todos los individuos en la población. Una hipótesis estadística es un enunciado general que puede ser expresada como una distribución de probabilidad sobre el espacio muestral, es decir, ésta determina una probabilidad para cada una de las posibles muestras.

De manera breve, la estadística es una disciplina que se encarga de, a través de muestras (cuantificadas como datos), describir el mundo. Y hay muchas cosas por describir: asociaciones, causalidad, realizar predicciones, establecer mecanismos de funcionamiento de objetos, etc. Es así como se establece su objetivo el cual de acuerdo con Wackerly et al. (2014) es:

realizar una inferencia sobre la población con base en la información contenida en una muestra de dicha población y proveer una medida asociada de qué tan buena es la inferencia.

---

<sup>1</sup>Traducción y subrayado de Rodrigo

Dentro de la definición previa y objetivos hay que destacar varios términos que son de importancia. La primera es la **población**, cualquier conjunto (no vacío) de objetos. Una **población** es lo más general posible, no necesariamente involucra personas o seres vivos. Algunos ejemplos de poblaciones incluyen: las personas que viven en Guatemala (si me interesa saber algo de los guatemaltecos en general), los árboles del Amazonas (si quiero saber cosas de ecología), los perros callejeros en Ciudad de México, los consumidores de una marca de cereal, los coches que transitan por Dubai, los granos de arena en una playa específica de Cancún, las células T dentro de los seres humanos o los metales pesados.

Más relevante que la población (para nuestros propósitos) es la **población objetivo**. El conjunto de elementos que formarán parte del estudio. Definir la **población objetivo** es complicado en algunas situaciones; por ejemplo, si se desea saber si *los mexicanos* están a favor o en contra de legalizar la marihuana hay que establecer quiénes son *los mexicanos*. ¿Cuentan las personas con nacionalidad mexicana que residen en el extranjero? ¿Cuentan los menores de edad? ¿Qué pasa con los extranjeros que son residentes? De nuevo, la población objetivo no necesariamente son personas, es sólo aquello que nos interesa medir.

Idealmente el estudio estadístico sería sobre la población objetivo. Por ejemplo, si nos interesa estudiar la evolución de los enfermos de VIH, la **población objetivo** serían los enfermos. Sin embargo, en el mundo real es imposible conseguir a toda la población objetivo (dentro de los enfermos, por ejemplo, están aquellos que aún no saben que tienen la enfermedad y no acudirían a nuestro estudio). La **población muestreada** resulta de esta dificultad. La **población muestreada** es el conjunto de elementos sobre los cuales se construyó la muestra para el análisis estadístico. En el caso de los enfermos de VIH la **población muestreada** podrían ser las personas que para una fecha específica habían sido diagnosticadas (y nos olvidamos de quienes desconocen su diagnóstico) o toda la población mexicana (y llevamos kits de diagnóstico con nosotros cuando diagnosticuemos). En encuestas de consumo, por ejemplo, usualmente no se muestrean zonas remotas o de muy bajos recursos por lo que la **población muestreada** no coincide con la **población objetivo** (todos los consumidores) sino que son sólo los consumidores de mayor poder adquisitivo. En encuestas de elecciones si bien la población objetivo son *todas las personas que voten el día de la elección*, como la mayoría se hacen *antes* de la elección (exceptuando las de salida) entonces se aproxima la definición de *votante* buscando incluir sólo aquellos que estén registrados en el padrón electoral o bien aquellos que al ser encuestados digan que *sí* van a votar. Aquí la **población muestreada** tampoco coincide con la objetivo.

Una **muestra** es un subconjunto de la población muestreada. Si la muestra coincide con la población muestreada (es decir, muestreaste a todo el mundo) se dice que es un **censo**. Si se tiene un censo se conoce TODA la población por lo que no es necesario hacer ningún análisis de inferencia (ya sabes todo de todos). Puedes realizar predicciones o descripciones. Ejemplos de censos son las encuestas de fin de cursos, las calificaciones de todo un grupo o el registro

de todas las compras de todas las personas en una tienda en línea.

**Ojo** No hay que confundir la definición de **muestra** con la definición estadística de **muestra aleatoria** (ver más adelante) la cual es un tipo muy específico de muestra obtenida bajo reglas restrictivas.

Finalmente hay que definir **inferencia**, el propósito de estas notas. Para ello usaremos el ejemplo y una versión adaptada de la definición de Boghossian (2014). Considera que sabes dos verdades:

1. Llovió anoche
2. Cuando llueve el suelo se moja

por lo que esta mañana *inferes* que el suelo estará mojado y sales de tu casa con botas y no con chancas. El proceso de *inferir* parece una consecuencia lógica de las premisas 1 y 2 pero no lo es exactamente: hoy es otro día y si hizo suficiente calor en la noche el agua pudo haberse evaporado del suelo. De ahí que definamos inferencia como:

Realizar un juicio el cual se explica a partir de premisas que suponemos verdaderas.

En particular **la inferencia estadística** será la rama de la estadística cuyo propósito es

Realizar juicios probabilísticos a partir de datos que suponemos verdaderos.

Aquí es necesario desglosar un poco la definición:

- Se habla de **juicios probabilísticos** pues nuestros juicios nunca van a ser tan certeros como **el suelo está mojado**. Más bien van a ser del estilo *hay una probabilidad muy alta de que el suelo esté mojado o nueve de cada diez veces el suelo estará mojado*.
- La **suposición de verdad** de los datos es muy relevante. Imagina el siguiente experimento: tu amiga borracha durante una fiesta se le ocurre que, de la nada, desarrolló poderes de psíquica y puede adivinar el futuro resultado de una moneda (cara o cruz). Tiras una moneda diez veces y todas las veces tu amiga hace una predicción correcta. *Considerando los datos como verdad concluirías que tu amiga es psíquica*. Una observación a profundidad de la moneda quizá te revele que es una moneda truqueada que siempre cae en cara. En ese momento cambiarías la **suposición de verdad** de los datos y la inferencia de que tu amiga es psíquica.

A lo largo de este libro aprenderemos lo básico para realizar inferencias estadísticas: observar datos y suponer verdades a partir de ellos. Tristemente la estadística nunca nos va a poder dar la verdad absoluta pero, si lo hacemos bien, es quizá lo más cerca que podamos estar de ella.

## 1.2 Modelos

La estadística funciona a partir de la construcción de **modelos**. Estos pretenden ser una forma de describir el mundo mediante teoría de la probabilidad y lo que se busca es utilizar dicha teoría para realizar inferencias. Estos modelos teóricos representan la forma en la que suponemos funciona la población. Para propósitos de estas notas diremos que los modelos viven *en el mundo de los modelos* o *mundo de las ideas*. Los datos observados, para poder distinguirlos, viven en *el mundo real*. Muchos de los modelos (no todos) se componen de **parámetros** que requieren para poder funcionar los cuales son estimados mediante **estadísticos** que se construyen a partir de los datos.

Para nuestros propósitos, los modelos que usaremos siempre construirán una población de la siguiente forma:

### 1.2.0.1 Población

Una población es un conjunto no vacío de variables (o vectores) aleatorias.

$$\mathcal{X} = \{X_1, X_2, \dots\}$$

Una población no necesariamente es finita. Por ejemplo, si nos interesa saber el tiempo que tarda un cliente en ser atendido en una llamada telefónica al banco quizá podemos suponer que la llamada telefónica tiene una duración descrita por un modelo Exponencial( $\lambda$ ). La población sería el conjunto infinito de todas las posibles llamadas telefónicas que se pueden realizar bajo este modelo. Por otro lado, un ejemplo finito de una población, son las caras de una moneda en un experimento donde busquemos, para una moneda específica, si caen más caras que cruces (cae más de un lado que del otro).

Una **muestra** es cualquier subconjunto (posiblemente infinito también) de la población.

### 1.2.0.2 Muestra

Una muestra  $\mathcal{M}$  de una población  $\mathcal{X}$  es cualquier subconjunto no vacío de  $\mathcal{X}$ . Es decir,  $\mathcal{M}$  es una muestra de  $\mathcal{X}$  si:

$$\mathcal{M} \subseteq \mathcal{X}$$

Pocas veces hablaremos de *muestras* de manera general y nos enfocaremos, sobre todo, en **muestras aleatorias**:



### 1.2.0.3 Muestra aleatoria

Una muestra aleatoria de tamaño  $n$ ,  $\mathcal{X}_{(n)}$ , de una población  $\mathcal{X}$  es un subconjunto finito (de tamaño  $n$ ), no vacío de  $\mathcal{X}$  donde sus elementos son **variables aleatorias independientes idénticamente distribuidas**. Es decir,  $\mathcal{X}_{(n)}$  es una muestra de  $\mathcal{X}$  si:

1. **Es una muestra:**  $\mathcal{X}_{(n)} \subseteq \mathcal{X}$ ,
2. **de tamaño  $n$ :** Cardinalidad $\left(\mathcal{X}_{(n)}\right) = n$ ,
3. **con variables independientes:** si  $X_i, X_j \in \mathcal{X}_{(n)}$  entonces  $\mathbb{P}(X_i \in A, X_j \in B) = \mathbb{P}(X_i \in A) \cdot \mathbb{P}(X_j \in B)$  para  $A, B$  conjuntos *medibles*,
4. **idénticamente distribuidas:** para todo  $i = 1, 2, \dots, n$  se tiene que  $X_i$  tiene función de distribución acumulada  $F_X$ .

El punto 4. de la definición pide que todas las variables descritas tengan la misma distribución. Por ejemplo, podemos pedir que todas sean exponenciales con el mismo parámetro o todas sean gamma con los mismos parámetros. El punto es que todas las variables aleatorias estén descritas con el mismo modelo y sean independientes entre sí.

El punto 3. de la definición puede escribirse de otras formas más amigables, por ejemplo, si suponemos que las variables aleatorias son continuas y tienen densidad  $f_X$  entonces la independencia puede escribirse como:

$$f_X(x_i, x_j) = f_X(x_i) \cdot f_X(x_j)$$

mientras que si son discretas con función de masa de probabilidad  $p_X$  tenemos:

$$p_X(x_i, x_j) = p_X(x_i) \cdot p_X(x_j)$$

La **muestra observada** así como la **muestra aleatoria observada** es el conjunto de *datos* que realmente viste. Mientras que la **muestra** y la **muestra aleatoria** viven *en el mundo de los modelos* y son variables aleatorias (constructos teóricos, como sabes, bastante complejos), la **muestra observada** es lo que se midió. Antes de dar la definición veamos un ejemplo con un dado.

### 1.2.0.4 Tiro de un dado

Se realiza un experimento para saber si un dado es justo (todos los lados tienen la misma probabilidad). Para ello se tira el dado  $n = 10$  veces y se registran los tiros: 2, 6, 1, 3, 3, 3, 5, 1, 3, 2.

*Mundo del modelo*

La población en este caso es el conjunto infinito de todos los posibles tiros del dado. De ese conjunto obtenemos una muestra aleatoria de tamaño

$n = 10$  (suponemos que los tiros son independientes entre sí) dada por:

$$X_{(n)} = \{X_1, X_2, X_3, \dots, X_{10}\}$$

donde  $X_i$  tiene la siguiente distribución:

$$\mathbb{P}(X_i = z) = \begin{cases} p_1 & \text{si } z = 1 \\ p_2 & \text{si } z = 2 \\ p_3 & \text{si } z = 3 \\ p_4 & \text{si } z = 4 \\ p_5 & \text{si } z = 5 \\ p_6 & \text{si } z = 6 \\ 0 & \text{en otro caso.} \end{cases}$$

donde  $\sum_{k=1}^n p_k = 1$  y  $p_k \geq 0$  para todo  $k$ . Lo que interesa en este estudio es *inferir* quiénes son las  $p_k$  para determinar si es más probable que caiga en un lado que en otro. Las  $p_k$  se conocen como parámetros.

*Mundo real*

Ya en la realidad en esos 10 tiros no observamos cualquier cosa, observamos valores específicos que hacen que **la muestra aleatoria observada** sea:

$$s_n = \{x_1, x_2, \dots, x_{10}\} = \{2, 6, 1, 3, 3, 3, 5, 1, 3, 2\}.$$

Por supuesto que repitiendo el experimento (volviendo a tirar el dado 10 veces) lo más probable es que **la muestra aleatoria observada** cambie (y veamos otros números) pero el modelo, reflejado en la **muestra aleatoria** (teórica), permanezca inmutable. Una forma de estimar las probabilidades podría ser mediante proporciones y calcular, por ejemplo, la probabilidad de que aparezca 1 como:

$$\hat{p}_1 = \frac{\text{Veces que aparece 1}}{n} = \frac{2}{10}$$

En este caso,  $\hat{p}_1$  dado por  $\frac{2}{10}$  es un **estimador observado** de la verdadera probabilidad  $p_1$  que vive en el mundo de los modelos (y jamás podremos conocer)

Armados con el ejemplo anterior realicemos la definición de las muestras observadas:

#### 1.2.0.5 Muestra observada

Una muestra observada es una colección no vacía de valores codificados como números reales los cuales corresponden a realizaciones de una muestra

$\mathcal{X}$ . Usualmente la denotamos:

$$s = \{x_1, x_2, \dots\}$$

donde las  $x_i$  **NO SON VARIABLES ALEATORIAS** sino que son datos **fijos** ya observados.

#### 1.2.0.6 Muestra aleatoria observada

Una muestra aleatoria observada es una colección no vacía de tamaño  $n$  de valores codificados como números reales los cuales corresponden a realizaciones de una muestra aleatoria  $\mathcal{X}_{(n)}$ . En particular suponemos que  $x_1$  es el valor observado de la variable aleatoria  $X_1$ ,  $x_2$  es el valor observado de la variable aleatoria  $X_2$  y así sucesivamente. Generalmente la denotamos por:

$$s_{(n)} = \{x_1, x_2, \dots, x_n\}$$

donde las  $x_i$  **NO SON VARIABLES ALEATORIAS** sino que son datos **fijos** ya observados.

Veamos un segundo ejemplo:

#### 1.2.0.7 Cantidad de personas que llegan a una tienda

En muchos casos la llegada de personas se supone que sigue una distribución Poisson. En este caso nos interesa estimar el número promedio de personas por día que hay en una tienda de la cual se han medido las siguientes cantidades (por día). Suponemos que las llegadas son independientes entre sí (la cantidad de gente que llegó un día no influye en la cantidad que llegó el otro).

Día	Número de personas
1	50
2	45
3	60
4	65
5	55
6	40

*Mundo del modelo*

La población en este caso es el conjunto infinito de todas las posibles formas en que en un día pueden llegar personas. De ese conjunto obtenemos una muestra aleatoria de tamaño  $n = 6$  (suponemos que las observaciones son independientes entre sí) dada por:

$$X_{(n)} = \{X_1, X_2, X_3, X_4, X_5, X_6\}$$

donde las  $X_i \sim \text{Poisson}(\lambda)$  (todas con el mismo  $\lambda$ ). Recordamos que la media de una Poisson es  $\lambda$  por lo que el **parámetro** que nos interesa estimar es  $\lambda$ .

*Mundo real*

A partir de las 6 llegadas observadas construimos **la muestra aleatoria observada**:

$$s_n = \{x_1, x_2, x_3, x_4, x_5, x_6\} = \{50, 45, 60, 65, 55, 40\}.$$

Una forma de estimar la media  $\lambda$  es mediante el siguiente **estimador observado**:

$$\hat{\lambda} = \frac{1}{6} \sum_{i=1}^6 x_i = 52.5$$

Ojo, esto no significa que  $\lambda$  sea 52.5. Significa que nuestra hipótesis de quién es  $\lambda$  es 52.5 y que esperaríamos la próxima vez en la tienda 52 ó 53 personas. En el mundo real *quién sabe cuánto vale  $\lambda$* , nuestra hipótesis es que vale 52.5 pero eso no necesariamente es la realidad.

Como ya establecimos, muchas veces el modelo utiliza un **parámetro** el cual es desconocido. A partir de los datos construimos un **estimador observado** el cual es nuestra hipótesis del verdadero valor del parámetro. En general va a ser imposible que le atinemos al *verdadero* valor del parámetro pero la idea es que el **estimador observado** esté lo suficientemente cerca. En el ejemplo anterior nos gustaría, por ejemplo, que el verdadero parámetro quizá fuera  $\lambda = 52$  ó  $\lambda = 54$  pero nos sacaría mucho de onda que el parámetro real fuera  $\lambda = 1000000$ .

**1.2.0.8 Distribución paramétrica**

Una función de distribución acumulada es una **distribución paramétrica** con parámetro  $\vec{\theta}$  si dada una colección de distribuciones

$$\{F_{\vec{\theta}} | \theta \in \Theta\}$$

determinar  $\vec{\theta}$  determina la distribución. Es decir, la familia de distribuciones está indizada por  $\vec{\theta}$ . A  $\vec{\theta}$  se le conoce como el **parámetro** o **vector de parámetros**.

La definición anterior suena muy compleja sin embargo los ejemplos ya los conocemos.

#### 1.2.0.9 La normal

La distribución normal es una distribución paramétrica con

$$\vec{\theta} = (\mu, \sigma^2)^T$$

el vector de parámetros dado por la media y la varianza.

#### 1.2.0.10 La exponencial

La distribución exponencial es una distribución paramétrica con

$$\theta = \lambda$$

el parámetro que establece la tasa de la exponencial.

#### 1.2.0.11 La normal con varianza 1

La distribución normal con varianza 1 es una distribución paramétrica con

$$\theta = \mu$$

En este caso la varianza es conocida ( $\sigma^2 = 1$ ) pero la media no por eso sólo la media es el parámetro.

Podemos entonces definir un **estimador**:

#### 1.2.0.12 Estimador

Dada una distribución paramétrica  $F_\theta$  con parámetro  $\theta$  un estimador  $\hat{\theta}$  de  $\theta$  es una variable aleatoria que se construye como función de la muestra aleatoria:

$$\hat{\theta} : \mathcal{X}_{(n)} \rightarrow \Theta$$

Como  $\hat{\theta}$  es una función de la muestra aleatoria entonces puede representarse como:

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$$

Dado un conjunto de datos, el **estimador observado de  $\theta$**  es el estimador  $\hat{\theta}$  de  $\theta$  evaluado en los datos.

**1.2.0.13 Estimador observado**

Dada una distribución paramétrica  $F_\theta$  con parámetro  $\theta$  con estimador  $\hat{\theta}$  y datos observados  $s_{(n)} = \{x_1, x_2, \dots, x_n\}$  el **estimador observado** corresponde a la evaluación de  $\hat{\theta}$  en  $s_{(n)}$ ; es decir:

$$\hat{\theta}(x_1, x_2, \dots, x_n)$$

Veamos ejemplos para entender mejor cómo funciona esto.

**1.2.0.14 Tiros de una moneda**

Se tiene una moneda que cae más de un lado que del otro. Interesa estimar  $p$  la probabilidad de que caiga cruz. Para ello se toma una **muestra aleatoria** de 5 tiros de la moneda:

$$X_{(n)} = \{X_1, X_2, \dots, X_5\}$$

Suponemos que los tiros son independientes. El modelo entonces implicaría que

$$X_i \sim \text{Bernoulli}(p)$$

para cada  $i = 1, 2, \dots, 5$ . Si codificamos cruz como 1 y cara como 0, la **muestra aleatoria observada** es:

$$s_{(n)} = \{1, 1, 1, 0, 1\} = \{x_1, x_2, \dots, x_5\}$$

donde tuvimos tres cruces continuas, luego una cara y finalmente una cruz. Una opción de estimador observado sería contar la proporción de cruces haciendo:

$$\hat{\theta}(x_1, \dots, x_5) = \frac{1}{n} \sum_{k=1}^n x_i = \frac{4}{5}$$

de donde diríamos que nuestra hipótesis de cuánto vale el parámetro  $p$  es  $4/5$ . Por otro lado, el **estimador** teórico es:

$$\hat{\theta}(X_1, \dots, X_5) = \frac{1}{n} \sum_{k=1}^n X_i$$

el cual tiene una distribución de probabilidad sencilla pues  $\sum_{k=1}^n X_i \sim \text{Binomial}(n, p)$ . Particularmente podemos calcular su valor esperado, por

ejemplo,

$$\mathbb{E}[\hat{\theta}(X_1, \dots, X_5)] = \mathbb{E}\left[\frac{1}{n} \sum_{k=1}^n X_k\right] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] = \frac{1}{n} \sum_{k=1}^n p = \frac{1}{n} np = p$$

lo cual implica que el estimador, en promedio, devolvería el parámetro que nos interesa (esta propiedad se conoce como *ser insesgado* y lo veremos más adelante).

### 1.2.0.15 Error de medición de una app

Una app que se dedica a medir la altura de edificios mediante la toma de videos tiene un error de medición con distribución normal y cuya varianza es 1. Interesa determinar el error de medición promedio, el parámetro  $\mu$ . Para ello se toman videos y se miden edificios para obtener una colección de 7 errores de medición independientes en la siguiente **muestra aleatoria**:

$$X_{(n)} = \{X_1, X_2, \dots, X_5\}$$

El modelo es

$$X_i \sim \text{Normal}(\mu, 1)$$

para cada  $i = 1, 2, \dots, 7$ . Si los datos fueron:

Edificio	Error de medición
Bellas Artes	12.11
Torre Latinoamericana	40.54
Catedral Metropolitana	22.07
Palacio Nacional	15.22
Rectoría de la UNAM	45.18
Guerrero Chimalli	33.39
Estadio Azteca	41.76

la **muestra aleatoria observada** en este caso correspondió :

$$s_{(n)} = \{12.11, 40.54, 22.07, 15.22, 45.18, 33.39, 41.76\} = \{x_1, x_2, \dots, x_7\}$$

Una opción de estimador observado sería calcular la media muestral haciendo:

$$\hat{\theta}(x_1, \dots, x_7) = \frac{1}{n} \sum_{k=1}^n x_i = 30.03857$$

de donde diríamos que nuestra hipótesis de cuánto vale el parámetro  $\mu$  es 30.03857. Por otro lado, el **estimador** teórico es:

$$\hat{\theta}(X_1, \dots, X_7) = \frac{1}{n} \sum_{k=1}^n X_i$$

tiene una distribución de probabilidad sencilla pues sabemos que  $\sum_{k=1}^n X_i \sim \text{Normal}(\mu, \sigma^2)$ . Particularmente podemos calcular su valor esperado, por ejemplo,

$$\mathbb{E}[\hat{\theta}(X_1, \dots, X_7)] = \mathbb{E}\left[\frac{1}{n} \sum_{k=1}^n X_i\right] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_i] = \frac{1}{n} \sum_{k=1}^n \mu = \frac{1}{n} n\mu = \mu$$

lo cual implica que el estimador, en promedio, devolvería el parámetro que nos interesa (este también es *insesgado*).



# Bibliography

Boghossian, P. (2014). What is inference? *Philosophical studies*, 169(1):1–18.

Wackerly, D., Mendenhall, W., and Scheaffer, R. L. (2014). *Mathematical statistics with applications*. Cengage Learning.