Estimación Puntual

Introducción a la Estadística

Fiona Franco Churruarín
fionafch96@gmail.com

UTDT

Febrero 2022

De la muestra a la población

No olvidemos el *big picture*:

$$\begin{tikzcd}
\textbf{Población} \arrow[r, "\textit{muestreo}"] \arrow[d, leftrightarrow] & \textbf{Muestra} \arrow[d] \\
\textbf{Parámetro} & \textbf{Estadístico} \arrow[l, "\textit{inferencia}"']
\end{tikzcd}$$

De la muestra a la población

Al final del día nuestro objetivo es lograr hacer una declaración sobre alguna característica de la población.

Como observar a toda la población resulta imposible (costoso), nos centramos en una pequeña porción de ella, una *muestra*.

A partir de esta cantidad limitada de información, debemos elegir la mejor manera posible de procesarla para:

- dar una estimación (adivinanza) sobre el valor de una cantidad poblacional particular (parámetro); y
- aclarar que tan precisa es esa estimación.

Cualquier resumen de los datos de la muestra es llamado **estadístico**.

¿Que le pedimos a la muestra?

Vamos a considerar a cada unidad de la muestra como una variable aleatoria, y lo que llamaremos 'muestra aleatoria simple' cumple las siguientes propiedades:

- Todas las unidades provienen de la misma 'bolsa' (de la misma población). Si pensamos a todos los individuos de la muestra como un conjunto de variables aleatorias, este ítem quiere decir que todas tienen la misma distribución.
- Todas las unidades fueron elegidas de manera independiente, y sin priorizar una sobre otra. Esto quiere decir que todos los individuos en la población tienen la misma chance de ser elegidos para la muestra.

Bajo estas dos condiciones, si denotamos $X_1, X_2, ..., X_n$ a nuestra muestra de n mediciones, decimos que las $\{X_i\}_{i=1}^{n}$ son **independientes** e **idénticamente distribuídas**.

¿Qué es un estimador?

Por un momento llamemos a nuestro parámetro objetivo θ. Este valor podría ser el la altura media de las personas entre 18 y 70 años en Argentina, el efecto del gasto en publicidad sobre ventas de un determinado negocio, la probabilidad de que mañana llueva.

Decimos que un **estimador** es cualquier estrategia posible para intentar adivinar θ a partir de nuestra muestra. Lo denotamos con $\hat{\theta}$, nuestra herramienta. A diferencia de un estadístico, un estimador tiene como finalidad aprender sobre θ.

¿Qué es un estimador?

Suena a muchas posibilidades... Consideremos el caso de la media de la altura por simplicidad. Podríamos:

- Tirar los primeros 3 datos y usar la moda de los datos restantes.
- Usar toda la muestra y hacer un promedio ponderando por la edad de cada encuestado.
- Usar toda la muestra y decir que nuestra adivinanza es la media muestral.

Criterio

Para elegir entre alguna de todas las posibilidades hace falta un criterio, una regla de decisión tal que si yo proveo dos alternativas para estimar, esa regla me dice cuál es la preferida.

Hay una vasta literatura sobre lo que se conoce como 'funciones de pérdida' que cuantifican cuánto nos duele que $\hat{\theta}$ junto con un conjunto de datos particulares (estimación) no se parezca a θ (parámetro poblacional, el verdadero).

Nosotros enunciaremos algunas propiedades deseables que le pedimos a las herramientas (estimadores) que usaremos; y luego estudiaremos algunos estimadores y **por que** son utilizados en lugar de otros.

Primero veamos algunos ejemplos

Debido a sus propiedades, ciertos estadísticos son generalmente preferidos como la base para estimadores de la media poblacional (μ), la varianza poblacional (σ^2) o la proporción poblacional (p)

Parámetro	**Estimador**	**Estadístico**
μ	$\overline{X} = (\sum_i X_i)/n$	$\sum_i X_i$
σ^2	$S^2 = \frac{[\sum_i (X_i - \overline{X})^2]}{(n-1)}$	$\sum_i X_i$, $\sum_i X_i^2$
p	$\hat{p} = \frac{\text{nro éxitos}}{n}$	$\sum_i X_i$ o nro de éxitos

Aprendiendo $E(X) = \mu$

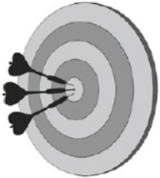
Uno pensaría '*y claro, si querés saber por donde anda la media poblacional de la altura de las personas, tomá una muestra y fijate cuánto es el promedio!*'.

Esta simple intuición es en este caso el mejor estimador para μ por dos motivos:

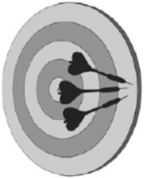
- En muchas muchas repeticiones del procedimiento (tomar una muestra, y calcular el promedio muestral), *en promedio*, el estimador acierta al parámetro μ. Es **insesgado**.
- En muchas muchas repeticiones del procedimiento, este estimador (herramienta) nos da estimaciones que son parecidas entre sí lo mas posible (resultados cercanos unos de otros para distintas muestras). Tiene poca **varianza**.

Es decir, el cañón apunta al blanco (es insesgado), y masomenos siempre dispara parecido (tiene poca dispersión, poca varianza).

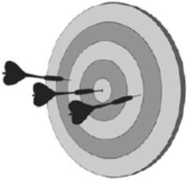
Sesgo y varianza de un estimador



(a) No bias, low variance



(b) Biased, low variance



(c) No bias, high variance

(d) Biased, high variance

Propiedades deseables - insesgamiento

Formalmente, decimos que un estimador $\hat{\theta}$ es **insesgado** para estimar θ si:

$$E(\hat{\theta}) = \theta$$

Se define el **sesgo** de un estimador como:

$$sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$$

La intuición viene de pensar en una herramienta que es utilizada muchas veces, con distintas posibles muestras. Si la usamos una y otra vez, nada garantiza que en un disparo particular acertemos o no al blanco. De todas formas, si es insesgado, *en promedio* de largo plazo acertaremos. Si es sesgado, el estimador erra persistentemente, no sólo por el azar de la muestra.

Varianza de un estimador

$$Var(\hat{\theta}) = E\left(\left[\hat{\theta} - E(\hat{\theta})\right]^2\right)$$

Un estimador tiene varianza... raro? Pero es así! Notar que un estimador $\hat{\theta}$ es una función de los datos de la muestra que nos toque.

Supongamos que en la población hay personas de 1.60, 1.70, 1.80, y hay exactamente un tercio de la poblacíon en cada categoría (para nada realista, pero ilustra el punto).

Varianza de un estimador

Vamos a tomar muestras de tamaño $n = 2$ y vamos a estimar la media poblacional $\mu = 1.70$ con el promedio muestral:

- Muestra 1: 1.60, 1.60. $\rightarrow$ Estimación: 1.60.
- Muestra 2: 1.70, 1.80. $\rightarrow$ Estimación: 1.75.
- Muestra 3: 1.60, 1.80. $\rightarrow$ Estimación: 1.70.

Es decir, depende la muestra que nos toque, la estimación puntual será distinta. La varianza del estimador cuantifica que tan distintos serán estos resultados en muchas muchas repeticiones. Una propiedad deseable de es que estos resultados sean lo mas parecidos posible.

Varianza de un estimador

Mas generalmente, la muestra aleatoria es un conjunto de variables aleatorias $\{X_i\}_{i=1}^n$, y el estimador es una combinación de estas variables aleatorias. En el caso anterior:

$$\hat{\mu} = \bar{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

Por lo tanto $\hat{\mu}$ también es una variable aleatoria. En este caso[1]

$$E(\bar{X}) = \mu$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Donde $\sigma^2 = Var(X_i)$, que son todas iguales pues las X_i son idénticamente distribuídas.

[1] Pueden intentar demostrarlo a partir de las propiedades de esperanza y varianza.

Consistencia de un estimador

Esta propiedad es mas abstracta pero al menos tan importante como las anteriores. Describe el comportamiento de un estimador cuando el tamaño de la muestra se vuelve grande, en particular cuando n tiende a infinito.

Decimos que un estimador $\hat{\theta}$ es **consistente** para θ si:

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \epsilon) = 1 \quad \forall \epsilon > 0$$

Esta definición es muy delicada, pero intuitivamente quiere decir lo siguiente. A medida que aumenta el tamaño de muestra, la herramienta que estamos usando se acerca cada vez mas a la verdad, en el sentido de que la probabilidad de que diste de la verdad muy muy poquito se hace 1.

Comparación entre estimadores

Un enfoque es anclar es sesgo que uno está dispuesto a tolerar (por lo general pedimos insesgamiento), y luego escoger el estimador de mínima varianza.

En algunos contextos puede resultar útil comparar distintos estimadores con sesgo positivo, y una medida que combina sesgo y varianza es el **error cuadrático medio**:

$$ECM(\hat{\theta}) = E\left(\left[\hat{\theta} - \theta\right]^2\right)$$

Puede demostrarse que:

$$ECM(\hat{\theta}) = sesgo(\hat{\theta})^2 + var(\hat{\theta})$$

Error cuadrático medio

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos posibles estimadores de θ, si:

$$ECM(\hat{\theta}_1) \leq ECM(\hat{\theta}_2)$$

Decimos que $\hat{\theta}_1$ es preferible a $\hat{\theta}_2$.

La función de ECM depende del valor verdadero θ, que desconocemos, por lo que las comparaciones pueden realizarse para algún valor dado de θ. Desafortunadamente no podemos encontrar un estimador ECM óptimo en el sentido de que sea mejor que cualquier otro posible estimador en términos de ECM para cualquier valor θ. **¿Por qué?** Piensen en un estimador $\hat{\theta} = c$ (igual a una constante).

Eficiencia

En el caso de tener diferentes estimadores insesgados, si:

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$$

Decimos que $\hat{\theta}_1$ es un estimador **eficiente** relativo a otro estimador insesgado $\hat{\theta}_2$ para estimar θ.

Para tener una idea mas directa de cómo se comparan dos estimadores insesgados en términos de su varianza, se define la eficiencia relativa como:

$$ER(\hat{\theta}_1, \hat{\theta}_2) = \frac{var(\hat{\theta}_2)}{var(\hat{\theta}_1)}$$

Eficiencia - Ejemplo

Consideremos que queremos estimar la media de la altura de las personas en Argentina, μ. Para ello tomaremos una muestra aleatoria simple de tamaño $n = 2$. Denotamos X_1 a la altura de la primer persona en la muestra y X_2 a la de la segunda. Así $E(X_1) = E(X_2) = \mu$, $var(X_1) = var(X_2) = \sigma^2$ y X_1, X_2 son variables aleatorias independientes.

Considere los siguientes posibles estimadores:

- $\hat{\theta}_1 = \frac{1}{3}X_1 + \frac{2}{3}X_2$
- $\hat{\theta}_2 = \frac{1}{2}X_1 + \frac{1}{2}X_2$
- $\hat{\theta}_3 = \frac{1}{4}X_1 + \frac{2}{4}X_2$

Preguntas:

- ¿Son todos insesgados?
- ¿Cuál es el mas eficiente de entre los insesgados?

Estimadores puntuales - Propiedades

Entonces, sea $\hat{\theta}$ un estimador puntual de θ, algunas propiedades deseables son:

- **Insesgamiento:** la esperanza de un estimador coincide con el valor real del parámetro a estimar, $E(\widehat{\theta}) = \theta$.
- **Eficiencia:** si un estimador insesgado posee menor varianza que otro estimador también insesgado, decimos que el primero es eficiente (relativo al segundo).
- **Consistencia:** cuando $n \to \infty$, entonces $\widehat{\theta} \xrightarrow{p} \theta$. Es decir, se hace muy probable que el estimador se "acerque" al parámetro.

Ley de los Grandes Números

Este resultado nos habla del comportamiento de un promedio a medida que aumenta la cantidad de observaciones que promediamos. Es un rasultado 'asintótico' en el sentido que describe qué pasa cuando $n \to \infty$.

Este teorema establece que un promedio simple de variables aleatorias independientes con la misma media se acerca más y más a la esperanza de estas variables aleatorias, a medida que n aumenta.

Ley de los Grandes Números: Si $X_1, X_2, \ldots$ son variables aleatorias independientes idénticamente distribuidas (IID), cada una con media $E(X) = \mu$. Entonces,

$$\overline{X} = \frac{\sum_i X_i}{n} \to \mu \quad \text{cuando} \quad n \to \infty$$

Ley de los Grandes Números

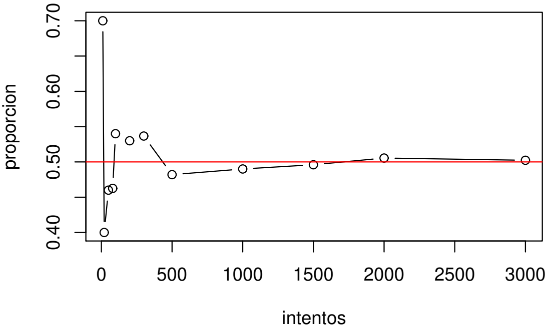
Simulamos un experimento en el que arrojamos una moneda.

Sabemos que la probabilidad de que salga cara es $1/2$. La siguiente tabla muestra cómo la **frecuencia relativa** de caras se acerca a la **probabilidad** de sacar cara medida que aumentamos la cantidad de repeticiones del experimento:

n	Caras	Cecas	F_{Cara}	F_{Ceca}
10	7	3	0.70	0.30
50	23	27	0.46	0.54
100	54	46	0.54	0.46
500	241	259	0.482	0.518
1000	490	510	0.490	0.510
3000	1507	1493	0.502	0.498

Ley de los Grandes Números

Ver `moneda.r`



Muestral vs Poblacional - Resumen

Todo concepto visto en estadística descriptiva tiene su correlato poblacional::

Muestral	**Poblacional**
Frecuencia relativa	Probabilidad
Histograma	Distribución de probabilidades
Media muestral ($\overline{X}$)	Media poblacional (μ)
Varianza muestral (S^2)	Varianza poblacional (σ^2)
Covarianza muestral (S_{XY})	Covarianza (COV(X,Y) o σ_{XY})
Coeficiente de correlación (r_{XY})	Coeficiente de correlación (ρ)

Distribuciones en el muestreo

Así como vimos que un estimador tiene varianza, y que dependiendo la muestra que nos toque obtendríamos estimaciones distintas, también vamos a interesarnos en la *distribución* de ciertos estimadores.

Recordemos que un estimador es una variable aleatoria en si misma, ya que es una función de las X_i de la muestra. En ese sentido podemos computar (en algunos casos) toda la distribución de un estimador, y esto sera esencial para cuantificar que tan seguros estamos de nuestra estimación puntual.

Media muestral $\bar{X}$ - Muestra Normal

Vimos que $E(\bar{X}) = \mu$ y $var(\bar{X}) = \frac{\sigma^2}{n}$.

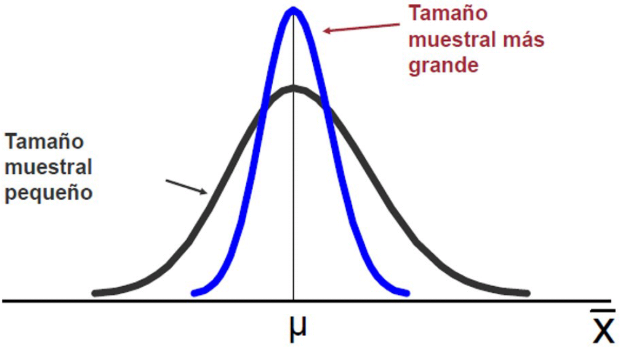
Un resultado importante, es que si tenemos $X_1, X_2, ..., X_n$ IID normales con media μ y varianza σ^2, la suma de estas variables aleatorias también será normal. Si las X_i son normales, $\bar{X}$ también lo será (dividir por n mantiene la normalidad).

Resultado: estandarizando, si la muestra proviene de una distribución normal,

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

Media muestral - Muestra Normal

Notar otra particularidad: la varianza del promedio muestral cae con el tamaño de muestra. Es decir que cuanto mas muestra tomemos aumentaremos la precisión del estimador.



Media muestral - Muestra no Normal

En general puede ocurrir que la población que estamos estudiando no siga una distribución normal...

Un famoso teorema matemático (el **Teorema Central del Límite**) establece que si el tamaño de la muestra es considerablemente grande, entonces la distribución de la media muestral (de n variables aleatorias IID) se aproxima a una distribución normal, **aunque la población no fuera normal**.

Este teorema es de esencial a la hora de realizar inferencia, ya que nos permitira tratar a la media muestral como una variable aleatoria que se distribuye *aproximadamente* normal, aunque la población no lo sea.

Teorema Central del Límite

A medida que n se incrementa, la distribución muestral de la media muestral estandarizada se aproxima a una distribución normal con media 0 y varianza 1.

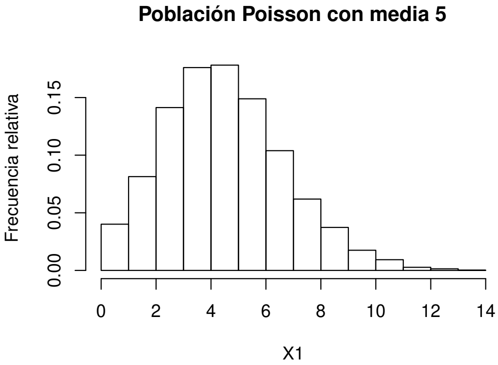
Entonces, para un n "*lo suficientemente grande*":

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \overset{aprox.}{\sim} \quad N(0,1)$$

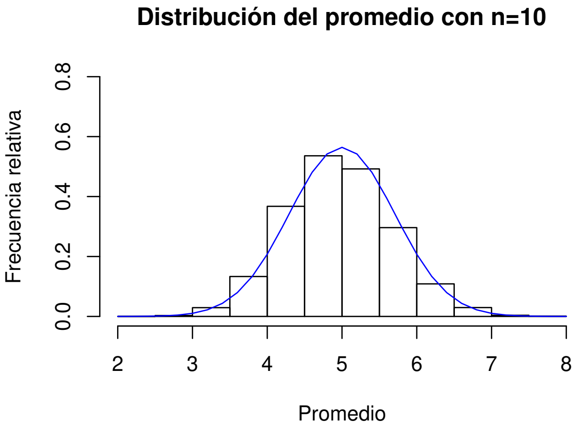
'Lo suficientemente grande' depende del contexto y genera controversia, a mayor n, mejor la aproximación.

Teorema Central del Límite

Ver `tcl.r`



Teorema Central del Límite



Varianza muestral (S^2)

Así como la media muestral tiene una distribución, la varianza muestral también.

Vimos que un estimador puntual de la varianza es:

$$S^2 = \frac{\sum_i (X_i - \overline{X})^2}{n-1}$$

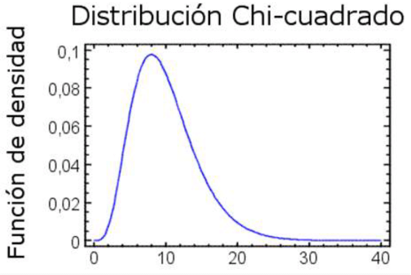
Usamos $n-1$ (y no n en el denominador) porque la media de su distribución muestral es la varianza poblacional, es decir, $E(S^2) = \sigma^2$ (es insesgado).

Distribución muestral de la varianza

Si la distribución poblacional es normal, se puede mostrar que:

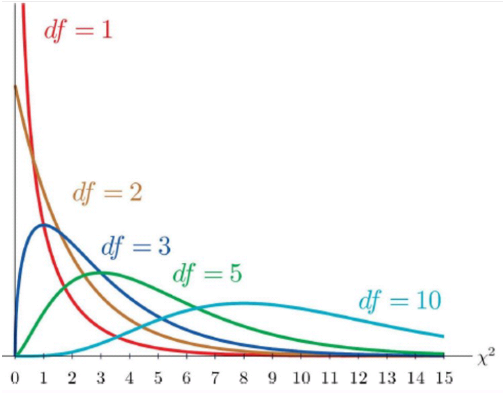
$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_i (X_i - \overline{X})^2}{\sigma^2}$$

tiene una distribución conocida como la **distribución χ^2 (chi-cuadrado) con $(n-1)$ grados de libertad.**



Distribución muestral de la varianza

Es una distribución con soporte en los reales positivos, asimétrica y se caracteriza por el número de grados de libertad:



Bibliografía y recursos útiles

- Newbold et al. (2013) - Cap. 6.

- Wheelan (2013) "The Central Limit Theorem. The Lebron James of Statistics", en Wheelan (ed.) *Naked Statistics*.
- Wooldridge - Apéndice C.
- Web: Galton Board Experiment