Taller de Reforzamiento

Introducción a Estadística

Sesión 2

Chamorro Rodriguez, Gianfranco David

E-mail: gianfranco.chamorror@gmail.com

Enero 2023

El presente documento sirve de apoyo para el taller de reforzamiento de Econometría, durante el cual se trabajarán conceptos que serán necesarios para iniciar un curso introductorio de pregrado o reforzar conceptos olvidados. Se brindará una perspectiva amable junto a material que permitara a los participantes abarcar temas mas complejos adelante.

---

1. Tipo de Variables

Una variable es cada una de las características o cualidades que poseen los individuos de una población, Por lo general consideramos dos tipos de variables :

- **Variable cualitativa**: Se refieren a características o cualidades que no pueden ser medidas o cuantificadas. dentro de esta hay dos tipos:

  Variable cualitativa nominal: Aquellas cualidades o características no cuantificables que no presentan orden.

  Variable cualitativa ordinal:Aquellas cualidades o características no cuantificables que si presentan orden.

- **Variable cuantitativa**: Se refieren a características o cualidades que pueden ser cuantificadas. dentro de esta hay dos tipos:

  Variable cuantitativa discreta: Es una variable que toma un valor finito de valores entre un intervalo de datos.

  Variable cuantitativa continua :Es una variable que toma un valor infinito de valores entre un intervalo de datos.

2. Probabilidades

La probabilidad es una medida de la certidumbre de que ocurra un evento. Es base para la estadística inferencial.

- **Probabilidad (Enfoque clásico)**: También conocida como probabilidad a priori.

  Es necesario definir dos conceptos: Frecuencia absoluta ($f(A)$) y relativa, el primero es el número de veces que ocurre un evento A al repetir n veces el experimento mientras que el segundo es la división del primero entre el número de veces realizado el experimento($f_r(A) = \dfrac{f(A)}{n}$) .

- **Probabilidad (Enfoque Frecuentista)**: Este enfoque nos indica que la probabilidad es la proporción de ocurrencias, frecuencias relativas, de un suceso o caso en una serie extensa de experimentos.[1]
  A nivel de frecuencia tenemos en cuenta estas 3 propiedades elementales :

$$0 \leq f_r(A) \leq 1$$

$$f_r(\emptyset) = \frac{0}{n} = 0$$

$$f_r(A \cup B) = f_r(A) + f_r(B)$$

  Con estas propiedades elementales, Kolmogorov definio la probabilidad de un experimento aleatorio con espacio muestral Ω.

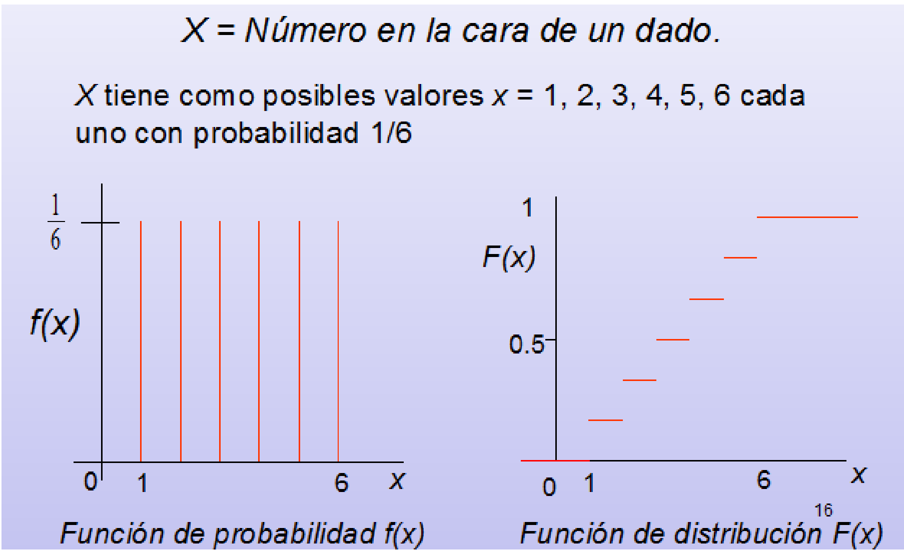
p(A) ≥ 0
$p(A \cup B) = p(A) + p(B)$; Si A y B son incompatibles
p(Ω) $= 1$

De estos tres axiomas :

p(A^c) $= 1 - p(A)$
p($\emptyset$) $= 0$
p(A$\cap B^c$) $= p(A) - p(A \cap B)$
p(A$\cup B$) $= p(A) + p(B) - p(A \cap B)$
A $\subseteq B \Longrightarrow p(A) \leq p(B)$
A $\subseteq \Omega \Longrightarrow p(A) \leq 1$

2.1. Distribuciones de Probabilidad

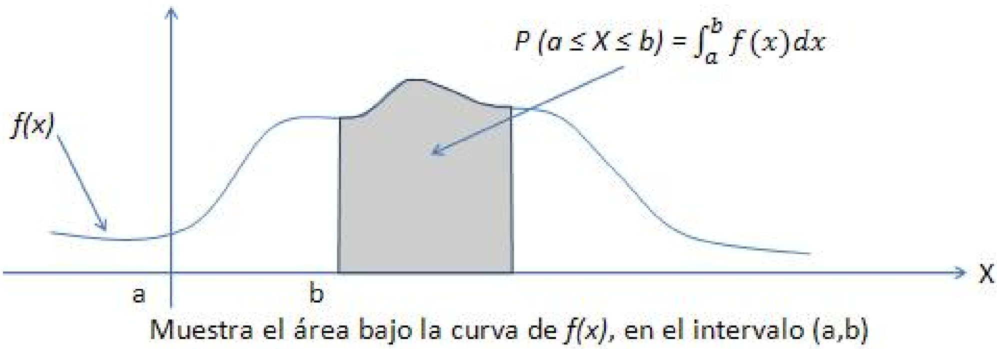
- **Función de probabilidad de variable aleatoria discreta**: Es una función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra, estando definida sobre el conjunto de todos de los posibles valores de la variable aleatoria.
- **Función de distribución acumulativa de una variable aleatoria discreta**: Nos brinda la probabilidad de que una variable aleatoria sea menor o igual a un valor dado.



[1] $\lim_{N\to\infty} \frac{n}{N}$

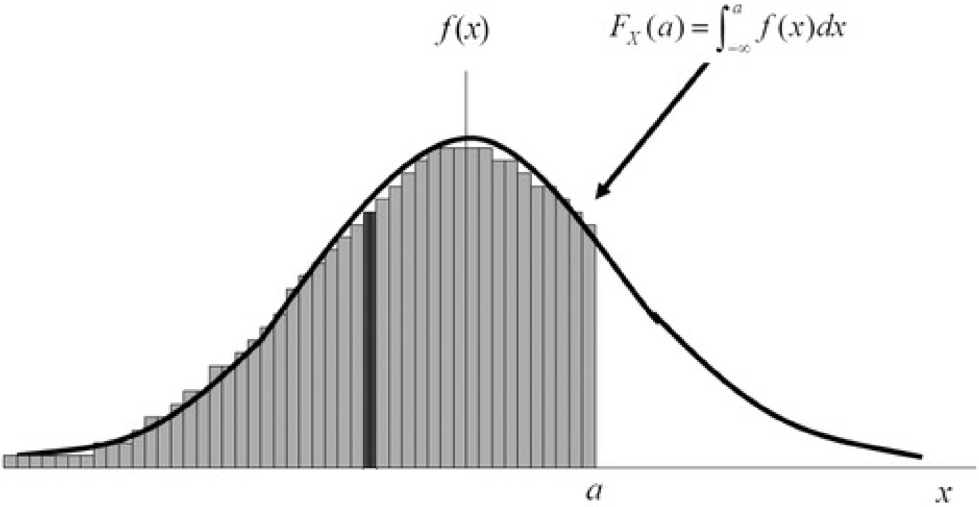
- **Función de densidad de probabilidad de variable aleatoria continua**:

  También conocida como densidad de una variable aleatoria continua, nos indica la probabilidad de que la variable tome un valor dentro de un rango determinado $f(x) = P(a \leq X \leq b)$ .La probabilidad de que la variable aleatoria caiga en una región específica del espacio de posibilidades estará dada por la integral de la densidad de esta variable entre uno y otro límite de dicha región.



- **Función acumulada de probabilidad de variable aleatoria continua**:

  Nos da la probabilidad de que la variable aleatoria sea menor o igual que dicho valor x. La función de distribución acumulativa de la variable aleatoria X es la función F(x) = P(X $\leq x$).

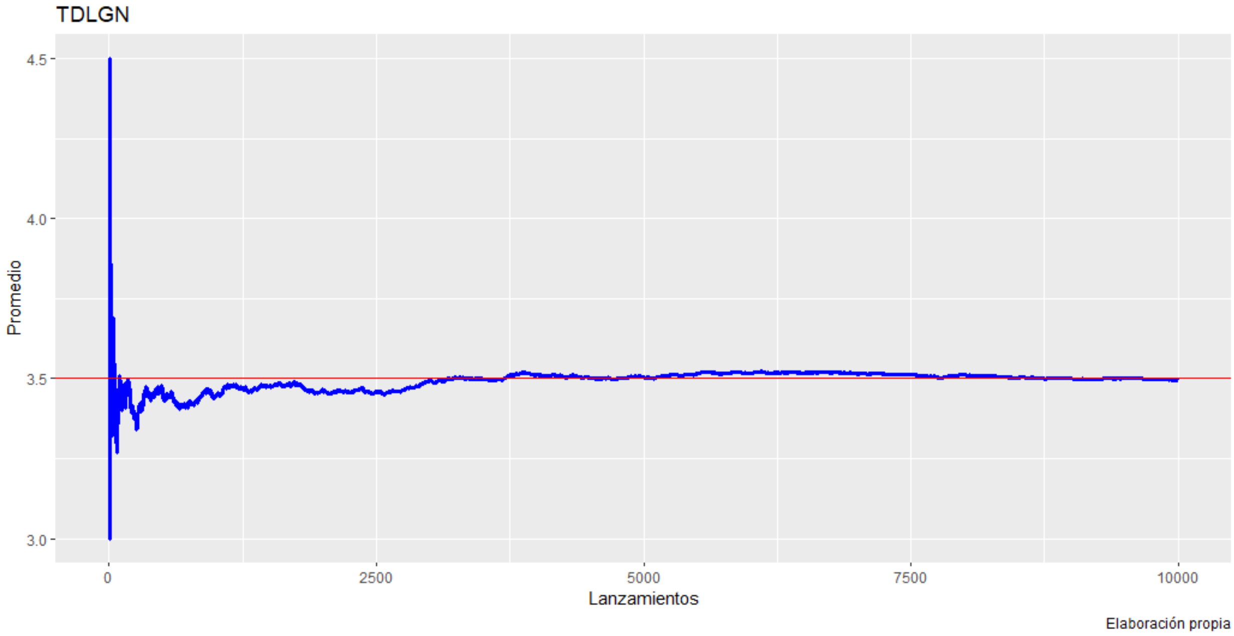


Ley de los Grandes Números

En la teoría de la probabilidad, bajo el término genérico de ley de los grandes números se engloban varios teoremas que describen el comportamiento del promedio de una sucesión de variables aleatorias conforme aumenta su número de ensayos.[2]. Es decir, la ley de los grandes números señala que si se lleva a cabo repetidas veces una misma prueba (por ejemplo, lanzar una moneda, tirar una ruleta,lanzar un dado, etc), la frecuencia con la que se repetirá un determinado suceso (que salga cara o sello, que salga el número 3 negro, o el promedio de los valores del resultado del dado, etc) se acercará a una constante. Esta será a su vez la probabilidad de que ocurra este evento.

[2] $\lim_{n\to\infty} \frac{f(A)}{n}$

Gráfico 1 : Ejemplo de lanzamiento de dado



Por la ley de Laplace la probabilidad es igual a la razón del número de casos favorables y los casos posibles siempre y cuando los sucesos elementales son equiprobables.Para ello es importante identificar los sucesos elementales de cada experimento, por ejemplo supongamos que lanzamos dos dados al mismo tiempo,definir la probabilidad de sacar un 2 o un 6 :

	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	(6,3)	10	11	12

P(X=2) $= \dfrac{CasosFavorables}{TotalCasos} = \dfrac{1}{36} = 0{,}0278$ $\qquad P(X=6) = \dfrac{CasosFavorables}{TotalCasos} = \dfrac{5}{36} = 0{,}1389$

3. Esperanza Matemática y Varianza

La esperanza matemática de una variable aleatoria X, es el número que expresa el valor medio del fenómeno que representa dicha variable.

$$\text{E[X]} = \begin{cases} \sum xf(x), \; si \; x \; es \; discreto \\ \\ \int xf(x)dx, \; si \; x \; es \; continua \end{cases}$$

Propiedades:

Sea X e Y , variables aleatorias y $a, b, c \in R$

- E[c]=c
- E[cX]=cE[X]
- Si X $\geq$ 0 entonces E[X]$\geq$ 0
- Si X $\leq$ Y entonces E[X] $\leq$ E[Y]
- Si $a < X < b$ entonces $a < E[X] < b$
- Si Y = a + bX, entonces $E[Y] = E[a + bX] = E[Y] = a + bE[X]$
- E[XY] = E[X] E[Y], Si X e Y son variables aleatorias independientes

La varianza de una variable aleatoria se expresa en términos de esperanza matemática: $Var[X] = E[(x-\mu)^2]$

$$\text{Var[X]} = \begin{cases} \sum (x-\mu)^2 f(x), \; si \; x \; es \; discreto \\ \\ \int (x-\mu)^2 f(x)dx, \; si \; x \; es \; continua \end{cases}$$

Propiedades:

Sea X e Y , variables aleatorias y $a \in R$

- Var[X]$\geq$ 0
- Var[a]=a
- Var[aX]=$a^2 Var[X]$
- Var[X$\pm$Y]= $Var[X] + Var[Y]$, Si X e Y son variables aleatorias independientes
- Var[XY]= $Var[X]Var[Y]$
- Var[X$\pm$Y]= $Var[X] + Var[Y] + 2COV[X, Y]$

Propiedades de la Covarianza:

- Cov[X,Y] = Cov[YX]
- Cov[X,Y] = E[XY] - E[X]E[Y]
- Cov[aX,Y] = aCov[X,Y]
- Cov[X,Y] = 0 , Si X e Y son variables aleatorias independientes
- Cov[$X_1 X_2, Y$] = Cov[X_1, Y] + Cov[X_2, Y]

El teorema del límite central es un teorema fundamental de probabilidad y estadística. El teorema describe la distribución de la media de una muestra aleatoria proveniente de una población con varianza finita. Cuando el tamaño de la muestra es lo suficientemente grande, la distribución de las medias sigue aproximadamente una distribución normal. El teorema se aplica independientemente de la forma de la distribución de la población. Muchos procedimientos estadísticos comunes requieren que los datos sean aproximadamente normales. El teorema de límite central le permite aplicar estos procedimientos útiles a poblaciones que son considerablemente no normales.

Gráfico 2 : Histograma de una variable

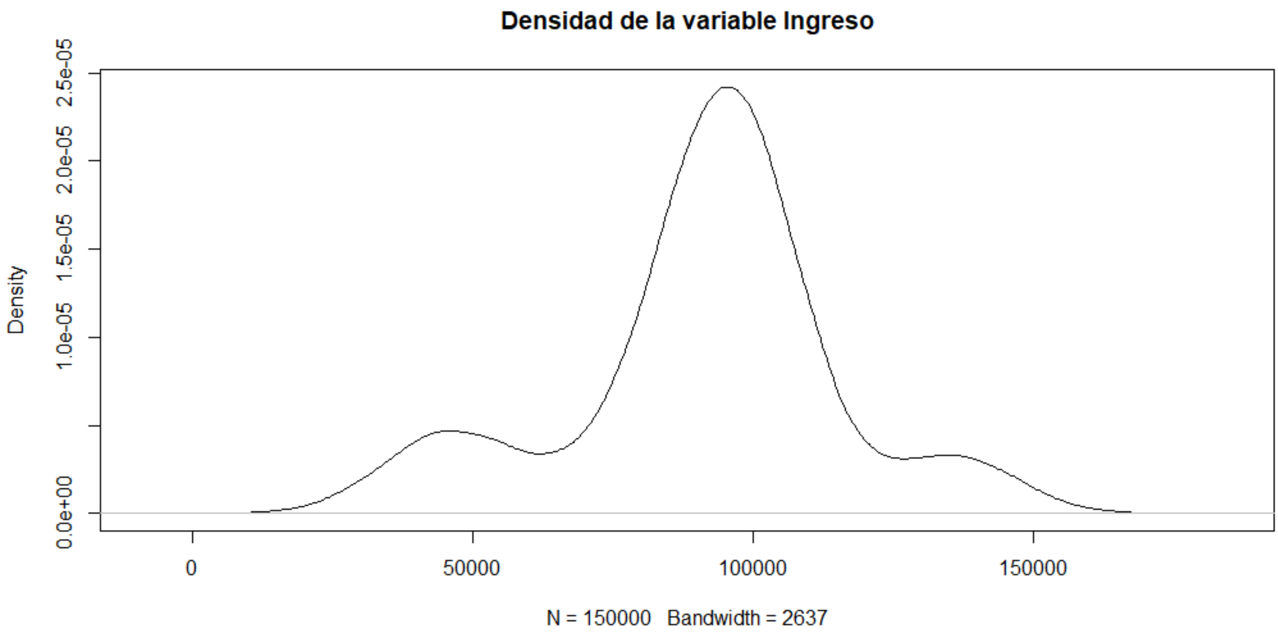


Gráfico 3 : Histograma del promedio de las muestras