Econometría I

Propiedades inferenciales

Carlos A. Yanes | Departamento de Economía | 2024-02-14

😄

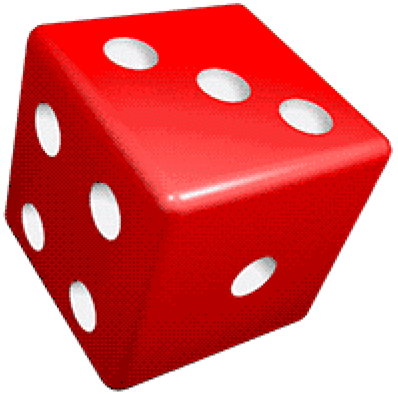
Preguntas de la sesión anterior?

Variables: Clasificación básica

> *Variable aleatoria*: Aquella que toma un **valor** numérico y su resultado esta determinado por un **Experimento** . Suelen escribirse con letras (X, Y, Z) y sus resultados se hacen con minúsculas (x, y, z).

Un ***ejemplo***: cuando se lanza un dado, X viene a ser el número de veces que lo lanzamos, entonces:

$X = \{1, 2, 3, \cdots, 6\}$ y (x) puede ser 4.



Variables: Clasificación básica

```
set.seed(123) # La opción de 1 es para una solo lanzamiento (resultado)
sample(1:6, 1)
```

```
#> [1] 3
```

Ya si quisieramos lanzar dos dados, tendriamos:

```
set.seed(123) # Ahora queremos ver dos resultados
sample(1:12, 2)
```

```
#> [1]  3 12
```

Variables: Clasificación básica

Para el caso del lanzamiento de una moneda:

$$X = \begin{cases} 1 & = & \text{Si es cara} \\ 0 & = & \text{Si es sello} \end{cases}$$

- La moneda tiene dos posibles **resultados**, *posiblemente encontramos la probabilidad de que la variable tome el valor de 1 si es cara con probabilidad 0.5*. Al igual que si sale sello. Ej:

$$P\left(X = 1\right) = 0.5 \text{ y por otro lado } P\left(X = 0\right) = 0.5$$

```
set.seed(123) # Garantizar el sorteo
sample(c("Cara", "Sello"), 1)
```

```
#> [1] "Cara"
```

Variables: Clasificación básica

- **Variables discretas**: Son aquellas que provienen de una naturaleza aleatoria o *resultado* aleatorio. Toman valores discreto de forma tal que $X : \{0, 1, 2, 3, \cdots, 245\}$

```r
# Definir la función para simular un sorteo de números discretos
simulacion <- function(n, min_valor, max_valor) {
  resultado <- sample(min_valor:max_valor, n, replace = TRUE)
  return(resultado)
}
# Vamos a simular un sorteo de 10 números discretos entre 1 y 100
resultado <- simulacion(10, 1, 100)
print(resultado)
```

```
#>  [1] 79 51 14 67 42 50 43 14 25 90
```

- **Variables continuas**: Aquellas que toman valores en un conjunto continuo de posibles valores. P.e.: $12.32; 114, 6$

```r
set.seed(123)
runif(n = 10, min = 15, max = 242)
```

```
#>  [1]  80.28010 193.94527 107.83776 215.44495 228.48607  25.34133 134.87995
#>  [8] 217.57912 140.17575 118.65154
```

Variables: Clasificación básica

Distribución

Distribución de probabilidad

Conceptos

- La **distribución de probabilidad** es una **relación** de todos los posibles valores que son posibles resultados de una variable.
- La **probabilidad** de los sucesos suele calcularse a partir de una distribución de probabilidad.
- La **distribución acumulada de probabilidad** resulta ser la probabilidad de que la **V.A** sea menor o igual a un valor concreto.

Probabilidad de ocurrencia		**Resultados**		
	0	1	2	3
Distribución de probabilidad	0.6	0.2	0.16	0.04
Distribución Acumulada	0.6	0.8	0.96	1

Recuerde 😮

- De la ecuación y frontera de probabilidad

$$p_i = P(X = x_i),\ i = 1, 2, 3, \ldots, k$$

  Donde,

$$0 \leq p_i \leq 1$$

- La función de **densidad** reúne toda la información acerca de los valores de X y sus probabilidades correspondientes:

$$f(x_i) = p_i,\ i = 1, 2, 3, \ldots, k$$

Distribución de una variable aleatoria discreta

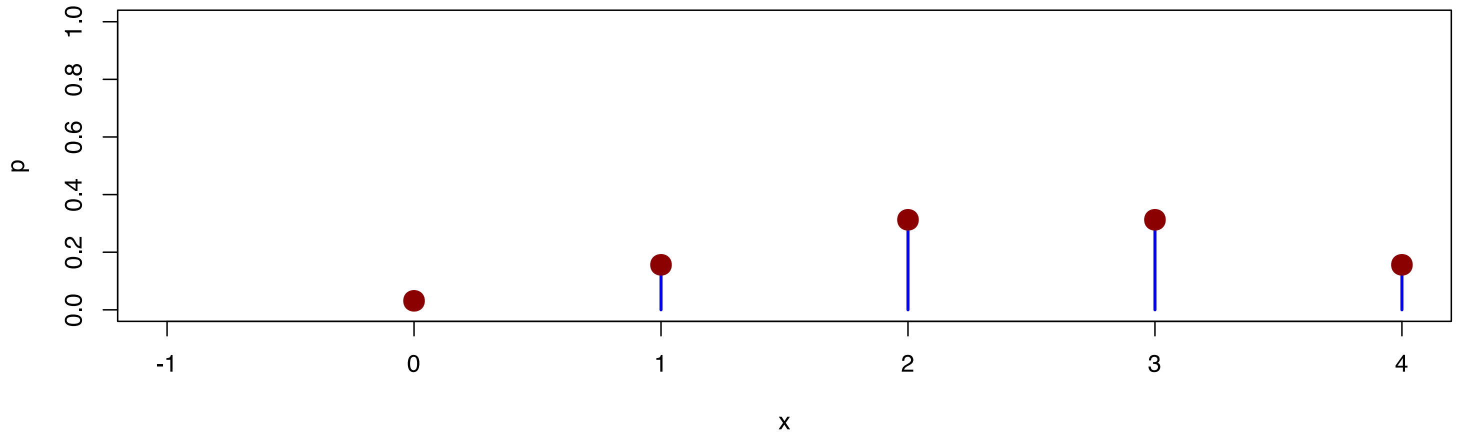
- Una **función de densidad discreta**, es $f(x)$ una *lista* de las probabilidades asociadas a diferentes realizaciones, x, que pueda tomar una variable discreta X:

$$P(W = w) : \begin{cases} \frac{1}{6} & si \quad w = 1 \\ \frac{1}{6} & si \quad w = 2 \\ \frac{1}{6} & si \quad w = 3 \\ \frac{1}{6} & si \quad w = 4 \\ \frac{1}{6} & si \quad w = 5 \\ \frac{1}{6} & si \quad w = 6 \\ 0 & si \quad ocurre\ otro\ evento \end{cases}$$

- El valor esperado del dado es:

$$E(W) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

Ejemplo de distribución discreta



Distribución de una variable aleatoria continua

- Una **función de densidad continua** es, $f(x)$, una función asociada a una variable continua X, de tal forma que:

$$\int_a^b f(x)dx = P(a \leq x \leq b)$$

Donde $f(x)$ debe cumplir:

- $f(x) \geq 0$.
- $\int_{-\infty}^{\infty} f(x)dx = 1$

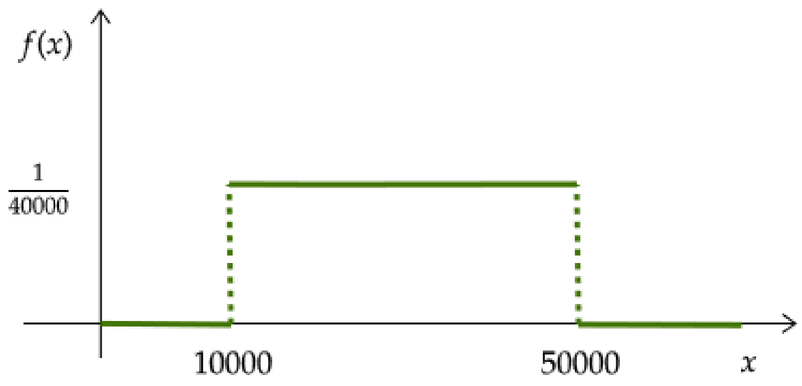
Esto permite conocer la probabilidad de que una V.A.C se encuentre en un intervalo.

Ejemplo: Distribución Continua

Una **empresa** después de haber realizado un estudio de la distribución del consumo de gasolina, encontró que:

$$f(x) = \begin{cases} \frac{1}{40000} & si \quad 10000 \leq x \leq 50000 \\ 0 & si \quad ocurre\ lo\ contrario \end{cases}$$

(x) viene a ser la **cantidad** de gasolina que consumen sus empleados en un mes.



Preguntas al público 🏆

Qué pasaría si quiero conocer:

1. La probabilidad que los empleados consuman en este mes exactamente 30000 litros de gasolina

2. Por lo menos 30000 litros

3. Entre 20000 y 30000 litros de gasolina

Para esto:

Resuelvo la integral de la **función**:

$$\int f(x)dx = \int \frac{1}{40000}dx = \frac{x}{40000} + C$$

Para la *pregunta 1* $\Rightarrow$

$$P(X = 30000) = P(30000 \leq x \leq 30000)$$

$$\int_{30000}^{30000} f(x)dx = 0$$

Es decir, la probabilidad que consuman 30 mil litros de gasolina es exactamente cero.

En R

```r
fp <- function(x){1/40000}
Vfp <- Vectorize(fp) # Se debe aplicar
e1 <- integrate(Vfp, lower = 30000, upper = 30000)$value
e1
```

```
#> [1] 0
```

Continuando:

Para la *pregunta 2* $\Rightarrow$

$$P(X \geq 30000) = 1 - P(10000 \leq x \leq 30000)$$

$$1 - \int_{10000}^{30000} f(x)dx = 1 - 0.5 = 0.5$$

La probabilidad es del 50% en este caso de que al menos consuman 30 mil litros.

De donde salió el **0.5**? 😱

$$\Big|_{10000}^{30000} \quad \frac{30000}{40000} - \frac{10000}{40000} = \frac{2}{4} = 0.5$$

😬

```
e2 <- 1 - integrate(Vfp, lower = 10000, upper = 30000)$value
e2
```

```
#> [1] 0.5
```

Para lo último

Para la *pregunta 3* $\Rightarrow$

$$P(20000 \leq X \leq 30000) = P(20000 \leq x \leq 30000)$$

$$\int_{20000}^{30000} f(x)dx = 0.25$$

La probabilidad es del 25% en este caso de que al consuman entre 20 mil y 30 mil litros de gasolina.

De donde:

$$\Big|_{20000}^{30000} \frac{30000}{40000} - \frac{20000}{40000} = \frac{1}{4} = 0.25$$

```r
e3 <- integrate(Vfp, lower = 20000, upper = 30000)$value
e3
```

```
#> [1] 0.25
```

Operadores

Operador: Sumatoria

> Simplifica de forma significativa el uso de términos en una expresión.

$$M_1 + M_2 + M_3 + M_4 + M_5 + M_6 + M_7$$

Entonces podemos escribir lo anterior como:

$$\sum_{i=1}^{7} M_i$$

Calcule el *resultado* de la siguiente expresión:

$$\sum_{j=0}^{2} \frac{1}{(j+1)(j+3)}$$

El desarrollo es sencillo:

$$\sum_{j=0}^{2} \frac{1}{(j+1)(j+3)} = \frac{1}{1 \cdot 3} + \frac{1}{2 \cdot 4} + \frac{1}{3 \cdot 5} = \frac{40 + 15 + 8}{120} = \frac{63}{120}$$

Propiedades Operador Sumatoria

Algunas de las principales:

1. $$\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i \quad \text{Prop. aditiva}$$

2. $$\sum_{i=1}^{n} ca_i = c\sum_{i=1}^{n} a_i \quad \text{Prop. de homogeneidad}$$

3. Una propiedad que es importante:

$$\sum_{i=1}^{n} c = n\ c\ ;\ \text{Prop. Constante}$$

Por ejemplo: $\sum_{i=1}^{4} 7 = 7 + 7 + 7 + 7 = 28$

Operador Productoria

> Es **análogo** al operador **sumatoria**. Se plantea como lo siguiente:

$$\prod_{i=1}^{n} a_i = a_1 \times a_2 \cdots a_n$$

Por ejemplo:

$$\prod_{i=2}^{4} (3i - 2) = 4 \times 7 \times 10 = 280$$

Pequeño ejemplo en R

Con el software, es fácil de implementar las operaciones anteriores ya sea con los comandos `sum` o `prod`.

- Halle la sumatoria y el producto de $a = 35$ donde $i : \{1, 2, 3, 4, 5\}$

$$\sum_{i=1}^{5} a_i, \text{ ademas de la } \prod_{i=1}^{5} a_i$$

```
a<-rep(35,5)
a
```

```
#> [1] 35 35 35 35 35
```

```
sum(a)
```

```
#> [1] 175
```

```
prod(a)
```

```
#> [1] 52521875
```

Otro ejemplo en R

- Para la función

$$\prod_{i=2}^{4}(3i-2)$$

- Halle el respectivo **resultado**

De forma manual es:

$$(3)(2)-2\times(3)(3)-2\times(3)(4)-2=280$$

- En R construiremos la formula haciendo uso de las **funciones**, para eso podemos decirle al programa que usaremos el *comando* `function`.

Funciones

> Una función **R**, permite obtener un resultado a partir de una formula anteriormente establecida y tiene varias partes:

Primero hagamos un ejemplo simple de establecer la formula de raíz cuadrada de un valor o número

```
raiz =
  function(x) {
    raiz_x = x^(1/2)  # formula que se va establecer
    return(raiz_x) ## Resultado
  }
```

A continuación la usamos con el valor de 144

```
raiz(144)
```

```
#> [1] 12
```

Funciones

Podemos ademas implementar otras opciones como las listas o la manera en como quiere que nos quede el resultado:

```r
raiz =
  function(x) {
    raiz_x = x^(1/2)  # formula que se va establecer
    return(list(valor=x, raiz_cuadrada=raiz_x))
  }
```

Testeamos

```r
raiz(144)
```

```
#> $valor
#> [1] 144
#>
#> $raiz_cuadrada
#> [1] 12
```

Funciones

Tambien se puede con el formato `tibble` de `tidyverse` y colocar como si se tratara de una matriz.

```
raiz =
  function(x) {
    raiz_x = x^(1/2)  # formula que se va establecer
    df = tibble(valor=x, raiz_cuadrada=raiz_x) # un dataframe
    return(df)
  }
```

Testeamos

```
raiz(144)
```

```
#> # A tibble: 1 × 2
#>   valor raiz_cuadrada
#>   <dbl>         <dbl>
#> 1   144            12
```

El resultado aparece como si fuera una tabla

Funciones

Para el caso de

$$\prod_{i=2}^{4}(3i-2)$$

```
caps<-function(x,y){a=3*(x:y)-2;m=prod(a);m}
# Los argumentos son X y Y, y es donde empieza y termina la productoria.
caps(2,4) # con la función
```

```
#> [1] 280
```

Note que para este caso usamos (;) para separar por operación

Media y varianza de una distribución

Media y varianza de una distribución

Para el caso general -esto es- para X_i, i.i.d para cualquier distribución:

$$\text{Media: } E(X) = \bar{X} = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu_x = \mu_x$$

Para la **Varianza** $\Rightarrow$

$$\begin{aligned}
\text{Varianza: } &= var(\bar{X}) \\
&= E\left[\bar{X} - E(\bar{X})\right]^2 \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu_x)\right]^2 \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} (X_i - \mu_x)\right]^2
\end{aligned}$$

🥲

Covarianza

Covarianza

> Una de las medidas que muestra como evolucionan dos variables.

$$cov(X,Y) = \sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)]$$

⚠ Lo que vendría a ser:

$$\sum_{i=1}^{k}\sum_{j=1}^{l}(x_j-\mu_x)(y_i-\mu_y)Pr(X=x_j, Y=y_i)$$

El signo de la **covarianza** es fundamental!!

⚠ Esto significa:

- **La dirección de la influencia de una variable sobre la otra.**

```r
x <- c(4.2, 4.6, 3.8, 4.1, 4.3)  # Datos de una asignatura
y <- c(2.1, 2.4, 2.8, 2.1, 3.2)  # Datos segunda asignatura
covarianza_xy <- cov(x, y) # Formula base
print(covarianza_xy)
```

```
#> [1] -0.0125
```

Correlación

> Debido a que la covarianza presenta dificultades en la interpretación dado el producto de las dispersiones de variables de Y y X. Surge el concepto o medida de dependencia entre Y y X y resuelve el problema de interpretación.

$$Corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(x)var(y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

⚗ La medida esta en un intervalo cerrado

$$-1 \leqslant corr(X,Y) \leqslant 1$$

```
correlacion_xy <- cor(x, y) # Formula correlación
print(correlacion_xy)
```

```
#> [1] -0.08998863
```

Coeficiente de Correlación

> Dado que la **Covarianza** depende de las unidades en que se miden las desviaciones, aparece un concepto mas estandarizado a la hora de medir relación y asociación entre un par de variables **aleatorias** y tener un concepto o interpretación mas "detallada".

> Es un **estadístico** que mide la relación (**signo**) y la fuerza (*magnitud*) de asociación lineal entre dos variables.

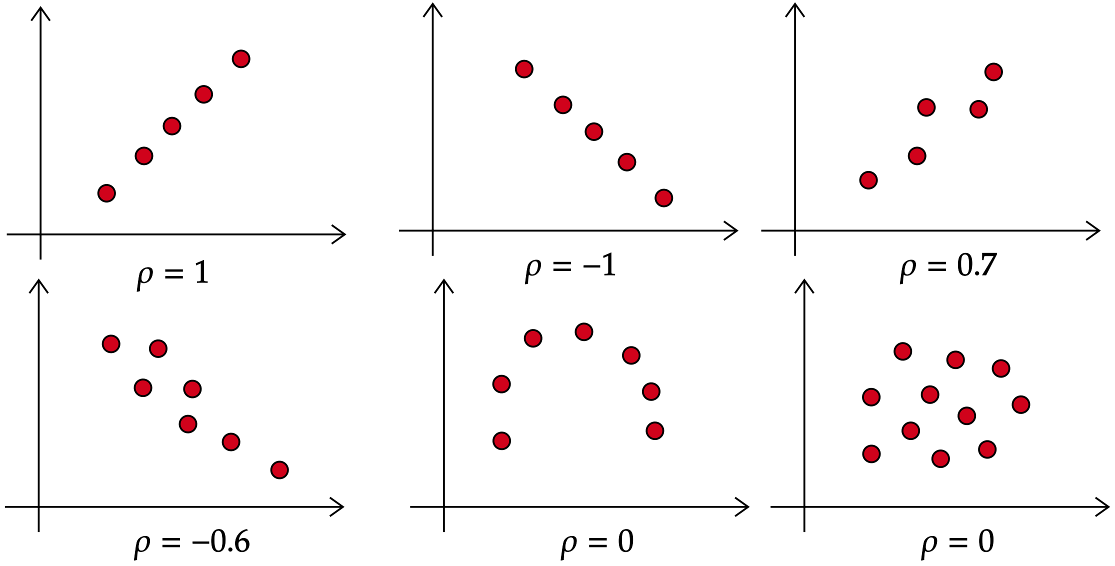
La formula puede ser *re-expresada* como:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Intepretación

El coeficiente (**r**) no tiene unidades y solo puede tomar valores entre -1 y 1, lo que es $-1 \leq r_{x,y} \leq 1$. **Su interpretación depende del signo y magnitud que nos arroja el estadístico**.

- Cuando $Corr(X, Y) = 1$, se dice que hay una asociación lineal perfecta y directa.
- Cuando $Corr(X, Y) = 0$, se dice que NO hay asociación.
- Cuando $Corr(X, Y) = -1$, se dice que hay una asociación lineal perfecta e indirecta.



Recuerde 😮

Propiedades en Varianza

La suma de varianzas X e Y toma la expresión:

$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y) = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}^2$$

Y dado el caso que las variables sean **independientes** la formula se reduce a:

$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y) = \sigma_x^2 + \sigma_y^2$$

Independencia lineal

⚗ Dos variables *aleatorias*, X y Y, se consideran independientes si y solamente si:

$$E[XY] = E[X]E[Y]$$

- Esto no implica que no existe relación entre ellas

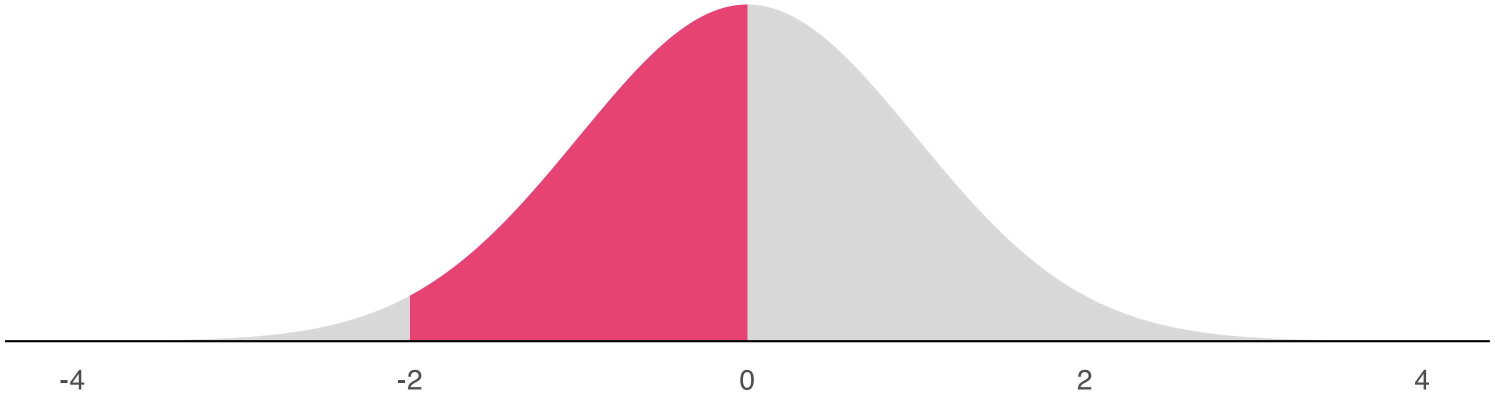
Nuevamente lo de distribución

Recordeis

Recordemos que utilizamos **funciones de densidad de probabilidad** (FDP) para describir la probabilidad de que una **variable aleatoria continua** tome un rango de valores, p.e: (El área total = 1). Estas *FDP* caracterizan las distribuciones de probabilidad, y las distribuciones más comunes/famosas/populares reciben nombres (*p.e*, normal, *t*, Weibull, Gamma, Poisson).

- La probabilidad de que una variable aleatoria normal estándar tome un valor entre -2 y 0:

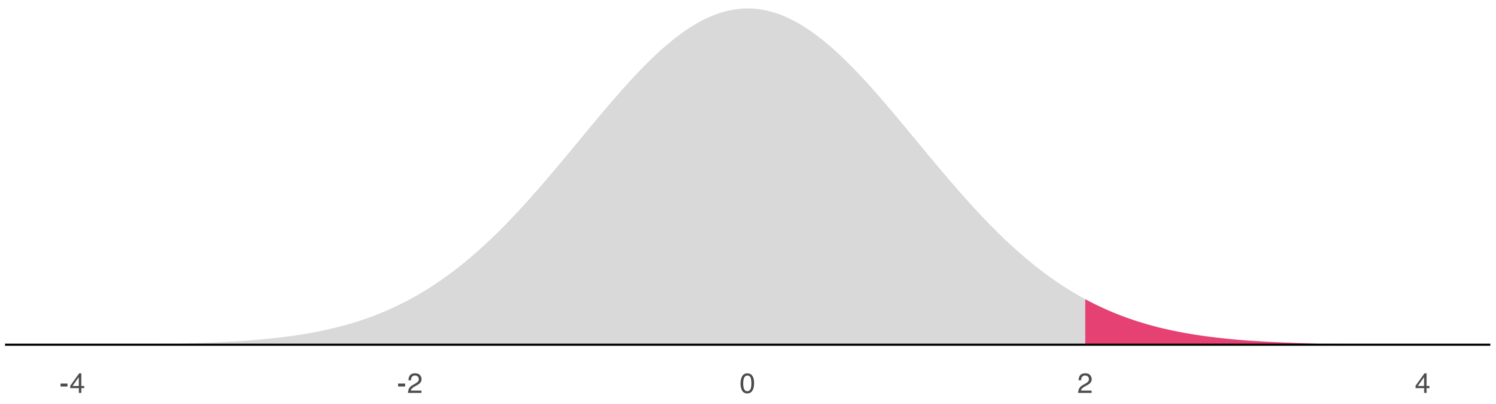
$$\mathrm{P}(-2 \leq X \leq 0) = 0.48$$



Recordeis

- Para el caso de hallar la probabilidad de que una variable aleatoria normal estándar tome un valor mayor a 2:

$$P(X > 2) = 0.023$$



En R

```r
prob <- 1- pnorm(2)
prob
```

```
#> [1] 0.02275013
```

Distribución Normal

> **Normal estándar**: Su forma concisa de expresión es $N(\mu, \sigma^2)$. Lo que significa que su media, $\mu = 0$ y su varianza $\sigma^2 = 1$. Para buscar probabilidades de una variable normal se debe **estandarizar** la variable en cuestión restando su **media** y luego dividir el resultado por la **desviación típica**.

Ejemplo: Suponga que desea calcular la probabilidad de que $Y \leq 2$? cuando $N \sim (1, 4)$.

- Para eso simplemente aplica:

$$Pr(Y \leq 2) = \frac{Y-1}{2} \Rightarrow \frac{2-1}{2} = \Phi(0.5) = 0.691$$

La formula de la normal es:

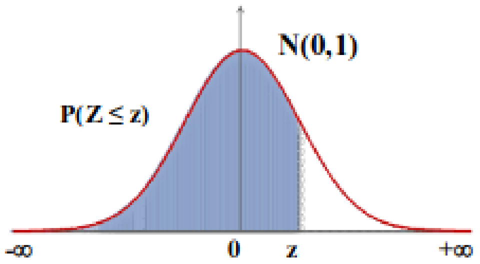
$$\text{Normal} = \frac{X-\mu}{\sigma}$$

- Donde X es el valor a *testear*; μ es la media de la **distribución** y σ la *desviación estandar*.

Mucho cuidado!!: El signo u operador de Φ significa que hay que ir a buscar el valor en la tabla de la Normal.

Ejemplo de valor en tabla normal

FUNCIÓN DE DISTRIBUCIÓN NORMAL N(0,1)



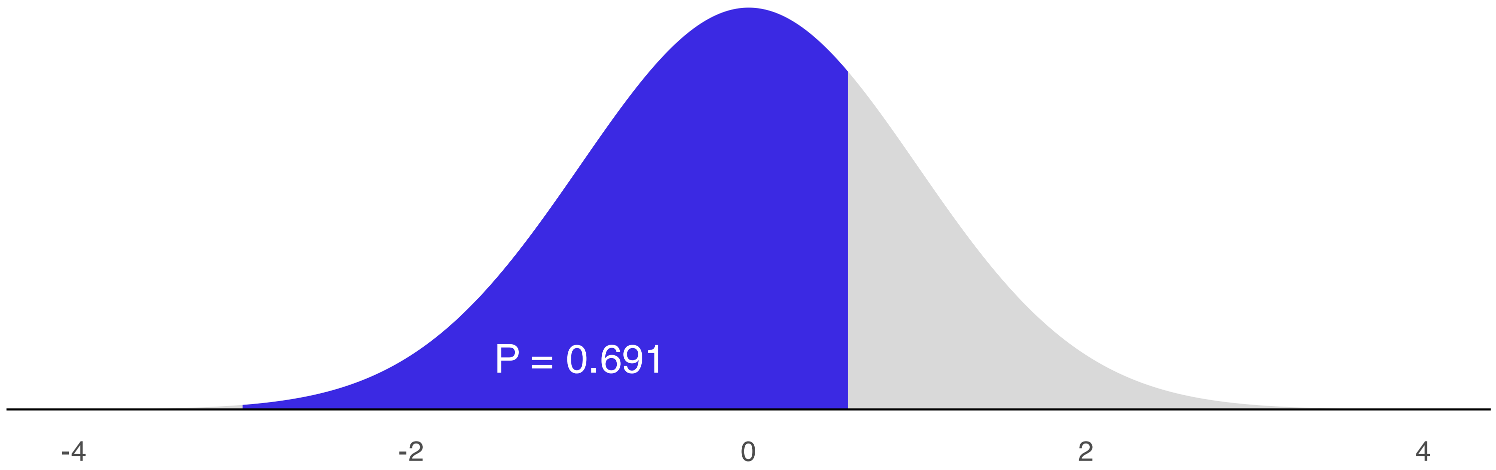
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852

En R es hacer uso del comando `pnorm()`, observe:

```
pnorm(0.5)
```

```
#> [1] 0.6914625
```

Grafico de nuestro ejemplo (normalizada)



La probabilidad de $P(x \leq 2) = P(Z \leq 0.5)$ con una media de 1 y varianza de 4 es del 69%.

Clave 📣

Sean X_1 y X_2 dos números tal que $X_1 < X_2$

- Es bueno saber que hay que usar lo siguiente normalmente:

$$Pr(Y \leq X_2) = Pr(Z \leq d_2) = \Phi(d_2)$$

- **Es posible que ocurra o le pregunten de esto**

$$Pr(Y \geq X_1) = Pr(Z \geq d_1) = 1 - \Phi(d_1)$$

- O también pase 🦸

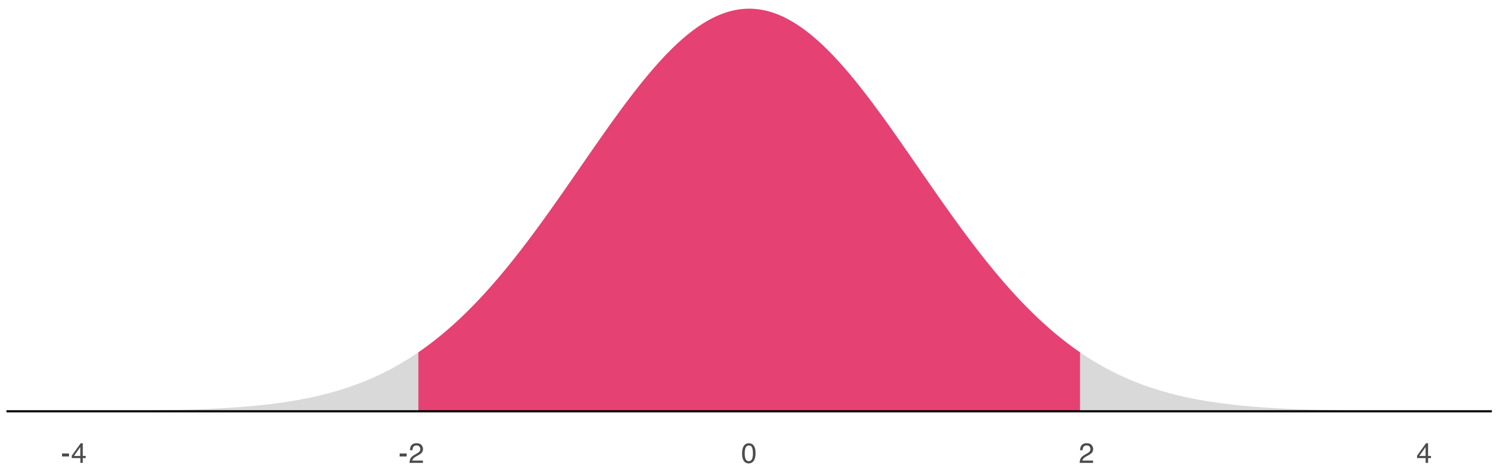
$$Pr(X_1 \leq Y \leq X_2) = Pr(d_1 \leq Z \leq d_2) = \Phi(d_2) - \Phi(d_1)$$

- No olvide que:

$$d = \frac{X_1 - \mu}{\sigma}$$

Mas ejemplos

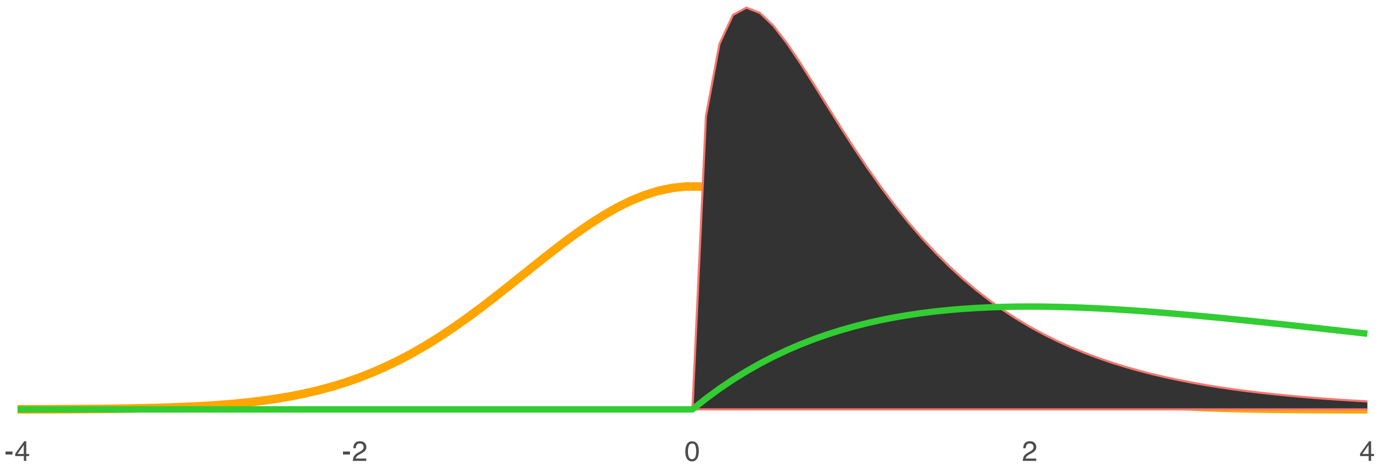
Ejemplo: Suponga que desea calcular la probabilidad de una variable aleatoria cuando $P(-1.96 \leq Y \leq 1.96) = 0.95$.




```r
probabilidad <- pnorm(1.96) - pnorm(-1.96)
probabilidad
```

```
#> [1] 0.9500042
```

Otras distribuciones



Las distintas distribuciones son importantes a la hora de hacer inferencia y pruebas de hipotesis y de **parámetros de los modelos**.

Y todo esto para qué? 😼

Estimador

Se considera como una variable aleatoria que posee las características de ser *Insesgado*, *Mínima Varianza* y *Eficiente*.

- Sea L una variable aleatoria en función de unas características tales como:

$$L = f(X_1, X_2, X_3, \cdots, X_n)$$

$$E(L) = \beta$$

Sesgo del estimador:

$$E(L) - \beta = 0$$

ℹ Nuestros estimadores muestrales, deben ser similares a nuestros estimadores poblacionales.

Estimación econométrica

Tenemos presente una ecuación econométrica:

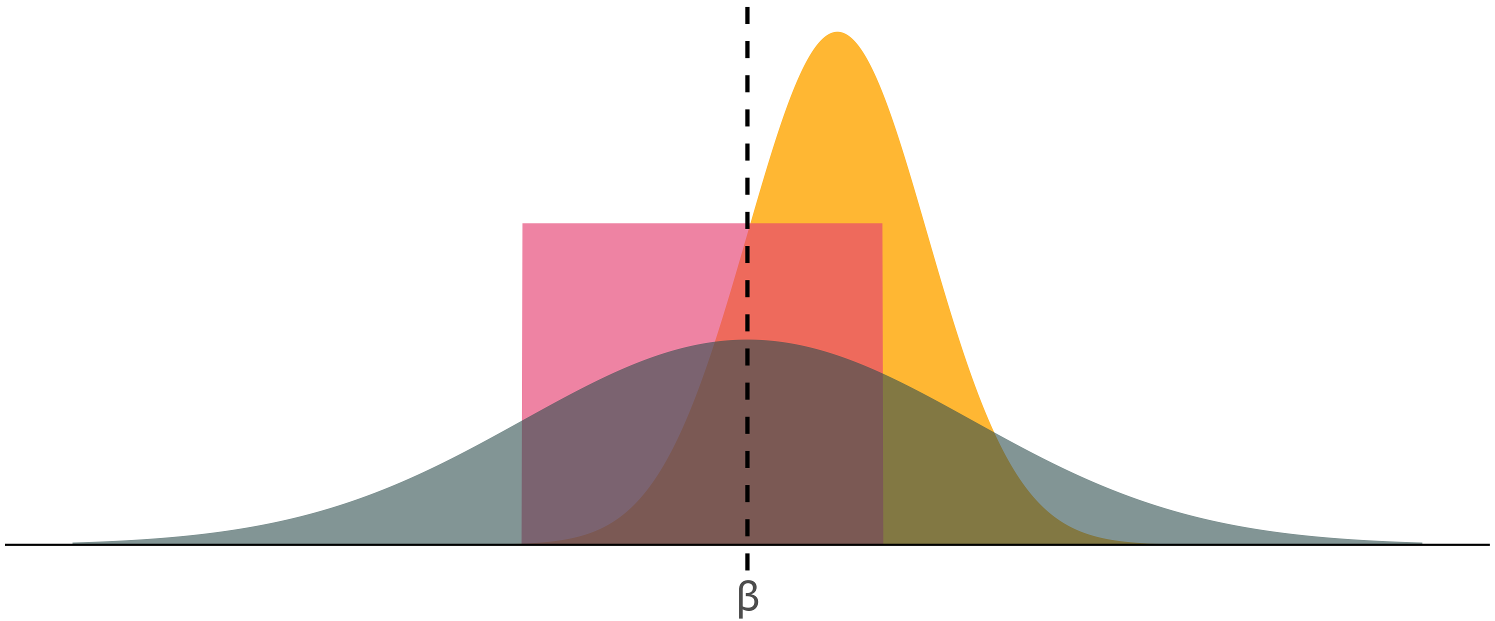
$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

- La ecuación tiene como objeto "modelar" la interacción de variables.
- La estadística es lo que permite validar estas estimaciones parámetricas.

Parámetro

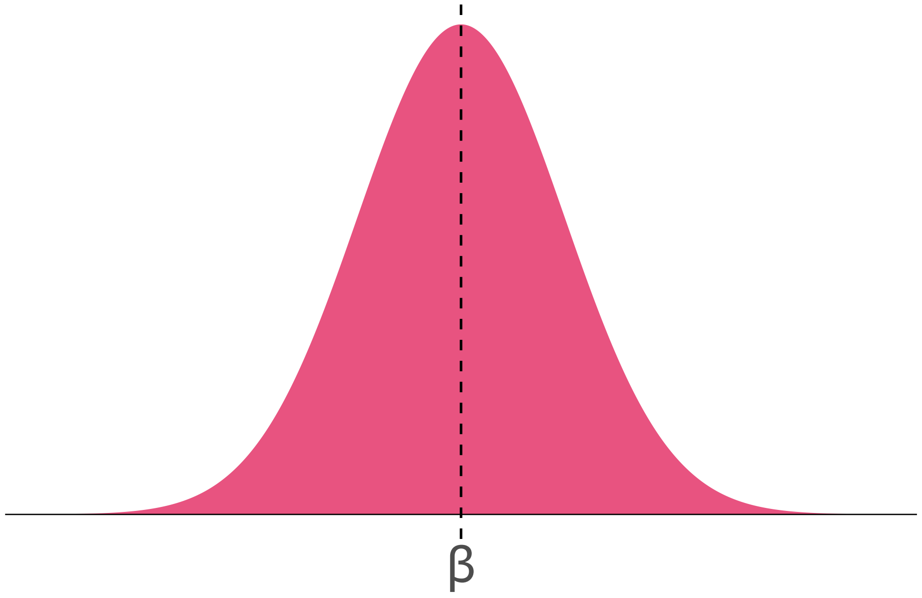
En nuestras ecuaciones lo tendremos como: $P = \{\beta, \alpha, \gamma, \theta, \cdots, \omega\}$ y lo tendremos como aquel valor de medida de cambio del efecto de una variable explicativa sobre la dependiente. *Función de datos que permite obtener estimaciones de los parámetros desconocidos*.

El estimador deseable

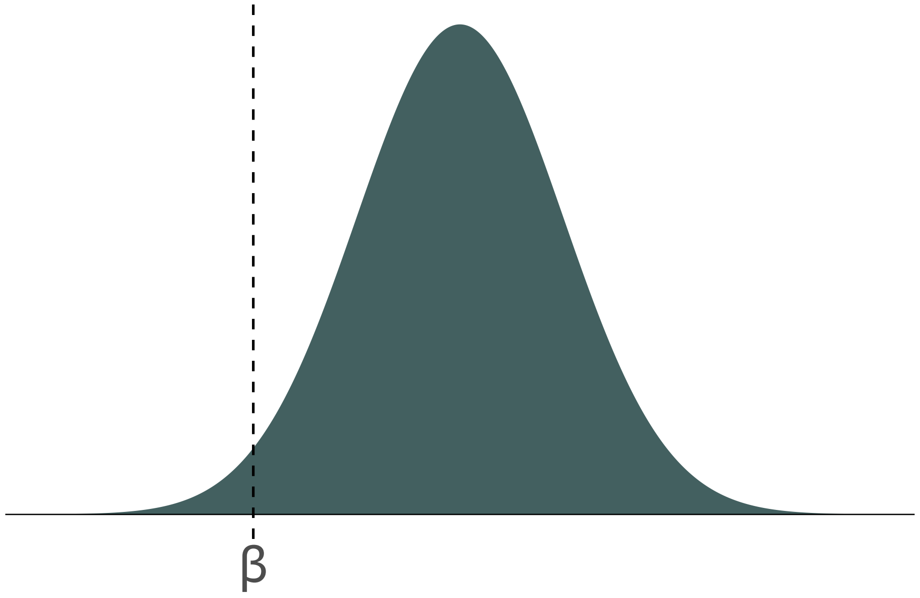


Sesgos del estimador

Estimator insesgado: $\boldsymbol{E}\left[\hat{\beta}\right] = \beta$



Estimator sesgado: $\boldsymbol{E}\left[\hat{\beta}\right] \neq \beta$



Y el asunto de la varianza en estimadores 👻

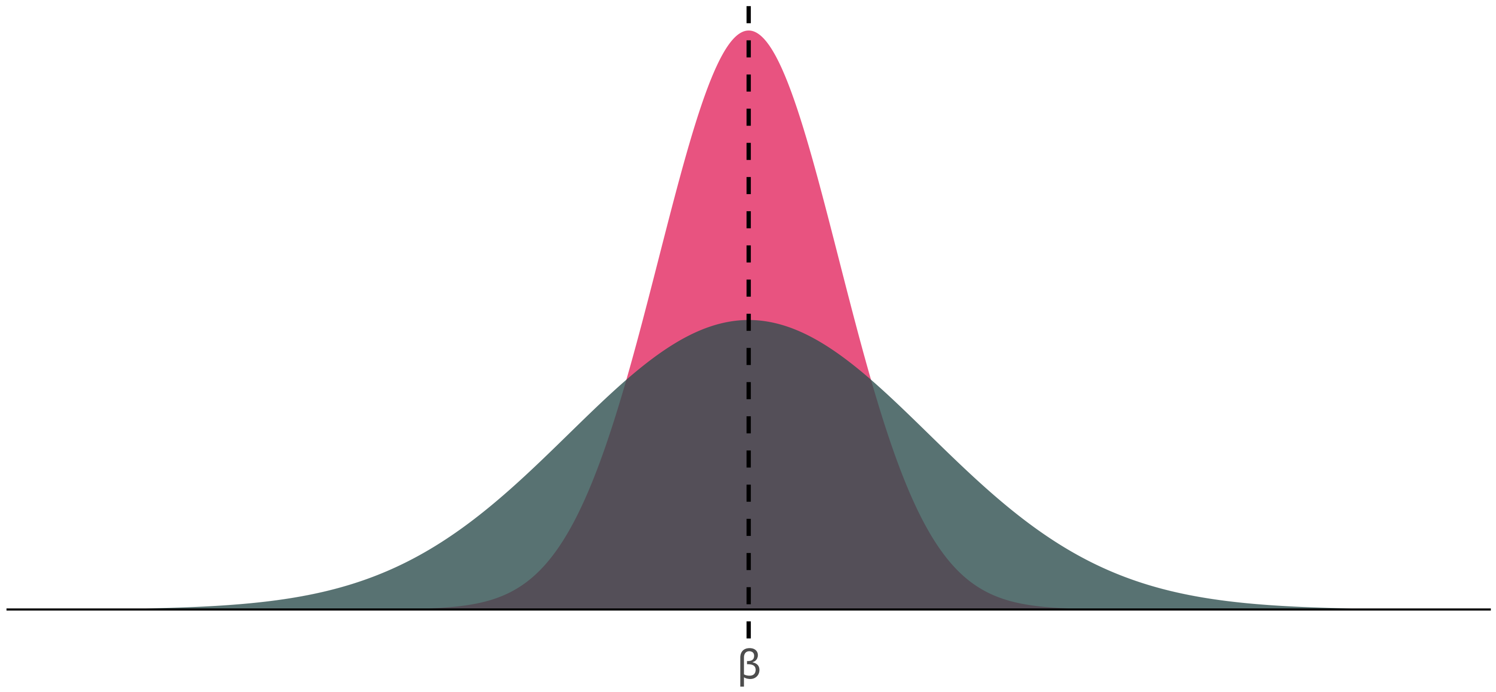
Mínima Varianza

Las tendencias centrales (medias) de las distribuciones en competencia no son lo único que importa. También nos importa la **varianza** de un estimador.

$$\text{Var} = (\hat{\beta}) = E[(\hat{\beta} - E[\hat{\beta}])^2]$$

Los estimadores de menor varianza significan que obtenemos estimaciones más cercanas a la media en cada muestra.

Mínima Varianza



Mínima Varianza

- Siempre hay una especie de **trade-off** entre sesgo y varianza.
- Para el caso de Econometría la preferencia ahonda la parte de insesgadez (consistencia).
- En otras partes… Ya va por el lado de lo que les sirve.. sin embargo hacen hincapie en la varianza.

Bibliografía

📕 Álvarez, R. A. R., Calvo, J. A. P., Torrado, C. A. M., & Mondragón, J. A. U. (2013). *Fundamentos de econometría intermedia: teoría y aplicaciones*. Universidad de los Andes.

📕 Stock, J. H., Watson, M. W., & Larrión, R. S. (2012). *Introducción a la Econometría*.

📕 Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Gracias por su atención!

Alguna pregunta adicional?

Carlos Andres Yanes Guerra

cayanes@uninorte.edu.co

keynes37