

Repaso

Econometría I

Pavel Solís

Universidad Anáhuac

2025

Econometría en una diapositiva

Planteamos un **modelo econométrico** para abordar una pregunta de interés

- Buscamos explicar una variable en términos de otras

Para estimar los **parámetros** del modelo, utilizamos variables aleatorias

- Los valores particulares de esas variables aleatorias se obtienen de los **datos**
- Esas variables aleatorias se llaman **estimadores**

En econometría, utilizamos los estimadores para:

- Estudiar sus propiedades
- Hacer **inferencia** sobre los parámetros del modelo

Esto con el fin de responder nuestra **pregunta de investigación**

Índice

Repaso de Matemáticas

Repaso de Probabilidad

Repaso de Estadística

Repaso de Matemáticas

Notación

Letras griegas

Letras griegas (minúsculas) comúnmente usadas:

α alfa

β beta

γ gama

δ delta

ε épsilon

θ teta

λ lambda

μ mu

π pi

ρ ro

σ sigma

ϕ fi

χ ji

Notación

Símbolos

Símbolos matemáticos comúnmente usados:

\neq no es igual

\approx aproximadamente igual

\forall para todo

$|$ tal que, condicional

Δ cambio, variación

\sum suma

\sim se distribuye

\Rightarrow implica

\nRightarrow no implica

\Leftrightarrow si y solo si

\rightarrow converge a, tiende a

$m \times 10^n, m \in n$ m por diez elevado a n

Suma

Si $\{x_i \mid i = 1, 2, \dots, n\}$ denota una secuencia de n números, la **suma** es

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

Propiedades: Para cualquier constante c ,

- $\sum_{i=1}^n c = nc$
- $\sum_{i=1}^n cx_i = c \sum_{i=1}^n x_i$
 - Si una variable en la suma no tiene subíndice, *sale* de la suma

Cuidado: Si $\{(x_i, y_i) \mid i = 1, 2, \dots, n\}$ es un conjunto de n pares de números, $\sum_{i=1}^n x_i y_i$ no se simplifica, $\sum_{i=1}^n \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i}$ y $\sum_{i=1}^n x_i^2 \neq (\sum_{i=1}^n x_i)^2$

Estadístico descriptivo

Un **estadístico descriptivo** resume numérica o gráficamente un conjunto de datos para describir sus características de forma concisa

Los estadísticos descriptivos incluyen:

- Medidas de tendencia central (media, mediana, moda)
- Medidas de variabilidad (rango, varianza, desviación estándar)
- Medidas de posición (cuartiles, percentiles)
- Gráficos (histograma, diagrama de dispersión, gráfico circular)

Medidas de Tendencia Central

Media o promedio muestral

Promedio muestral:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Propiedades:

- $\sum_{i=1}^n (x_i - \bar{x}) = 0$
- $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i (x_i - \bar{x}) = \sum_{i=1}^n x_i^2 - n(\bar{x})^2$

Medidas de Tendencia Central

Mediana muestral

Para obtener la mediana de una secuencia $\{x_i \mid i = 1, 2, \dots, n\}$ de n números, se ordenan de menor a mayor

- Si n es impar, se reporta el valor que está en el centro
- Si n es par, se reporta el promedio de los dos valores que están en el centro

Características:

- Media permite hacer operaciones
- Mediana es menos sensible a cambios en los valores extremos

Medidas de posición

Cuartiles

Los **cuartiles** son valores que dividen un conjunto de datos ordenados de forma ascendente en 4 partes iguales

- El primer cuartil (Q_1) indica el valor debajo del cual se encuentran 25% de las observaciones
- El segundo cuartil (Q_2 o mediana) divide los datos en dos partes iguales
- El tercer cuartil (Q_3) indica el valor debajo del cual se encuentran 75% de las observaciones

Medidas de posición

Percentiles

Los **percentiles** son valores que dividen un conjunto de datos ordenados de forma ascendente en 100 partes iguales

- Ej. El percentil 20 (P_{20}), es el valor debajo del cual se encuentran 20% de las observaciones
- $P_{25} = Q_1$
- $P_{50} = Q_2 = \text{Mediana}$
- $P_{75} = Q_3$

Proporciones y Porcentajes

Un **porcentaje** se obtiene al multiplicar una proporción por 100 (ej. 0.82 a 82%)

Una **proporción** es la forma decimal de un porcentaje (ej. 28% a 0.28)

Cambios:

- Cambio en cantidad (cambio absoluto): $x_1 - x_0 = \Delta x$ (ej. de 10 a 12, 2)
- Cambio proporcional (cambio relativo): $\frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x_0}$ (ej. de 10 a 12, 0.2)
- Cambio porcentual: $\% \Delta x = 100 \left(\frac{\Delta x}{x_0} \right)$ (ej. de 10 a 12, 20%)

Cuando x está en porcentaje, hay 2 formas de describir el cambio:

- Cambio en puntos porcentuales: Δx (ej. de 4% a 6%, 2 p.p.)
- Cambio porcentual: $\% \Delta x$ (ej. de 4% a 6%, 50%).

Funciones lineales

Una variable

y es una función lineal de x si

$$y = \beta_0 + \beta_1 x,$$

donde β_0 es el intercepto y β_1 es la pendiente

- La relación lineal implica que $\Delta y = \beta_1 \Delta x$
- $\Delta x = 1$ genera el mismo Δy sin importar el valor inicial
- β_1 es el **efecto marginal** (constante) de x sobre y

Una línea recta tiene la misma pendiente ($\Delta y / \Delta x$) en todos sus puntos

Funciones lineales

Dos variables

y es una función lineal de dos variables x_1 y x_2 si

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

donde β_0 es el intercepto, y β_1 y β_2 son las pendientes

- $y = \beta_0$ cuando $x_1 = 0$ y $x_2 = 0$
- La relación lineal implica que $\Delta y = \beta_1 \Delta x_1 + \beta_2 \Delta x_2$
- Si $\Delta x_2 = 0$, $\Delta y = \beta_1 \Delta x_1 \implies \beta_1 = \frac{\Delta y}{\Delta x_1}$ (pendiente en la dirección de x_1)
- Si $\Delta x_1 = 0$, $\Delta y = \beta_2 \Delta x_2 \implies \beta_2 = \frac{\Delta y}{\Delta x_2}$
- β_1 es el **efecto parcial** de x_1 sobre y (cómo cambia y con x_1 cuando x_2 fijo)
- β_2 es el efecto parcial de x_2 sobre y

Combinaciones lineales

Si c_1 y c_2 son 2 constantes y p y q son 2 variables, entonces

$$m = c_1 p + c_2 q$$

- m es una nueva variable
- m es una **combinación lineal perfecta** de p y q

Ejemplos:

- $c_1 = 1$ y $c_2 = 1$
- $c_1 = 1,000$ y $c_2 = 0$

Misma idea se puede extender a k constantes y k variables

Funciones No Lineales

En una función no lineal, el cambio en y para un cambio dado en x depende del valor inicial de x

- Permiten capturar rendimientos marginales decrecientes o crecientes
 - Insumo aumenta manteniendo otros fijos, producción adicional disminuye
 - Todos los insumos aumentan, producción aumenta en mayor proporción

Funciones no lineales comunmente usadas:

- Cuadrática
- Logaritmo
- Exponencial

Función cuadrática

y es una función cuadrática de x si

$$y = f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

Su derivada es:

$$f'(x) = \beta_1 + 2\beta_2 x^* = 0 \implies x^* = -\frac{\beta_1}{2\beta_2}$$

Si $\beta_1 > 0$ y $\beta_2 < 0$,

- x^* es un máximo
- El efecto marginal de x sobre y es decreciente
 - Efecto puede ser negativo: Para $x > x^*$, incrementos en x reducen y

Logaritmo natural

La función logaritmo natural:

$$y = \ln(x) = \log(x), \quad x > 0$$

- Muestra rendimientos marginales decrecientes pero el efecto nunca es negativo (pendiente $\rightarrow 0$)
- Es útil para hacer aproximaciones (ej. rendimientos accionarios)
 - $\log(1+x) \approx x$ cuando $x \approx 0$
 - $\log(x) \approx x - 1$ cuando $x \approx 1$

La diferencia de logaritmos se usa para aproximar cambios cuando $\frac{x_1}{x_0} \approx 1$

- $\Delta \log(x) = \log(x_1) - \log(x_0) = \log\left(\frac{x_1}{x_0}\right) \approx \frac{x_1}{x_0} - 1 = \frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x_0}$
- $\% \Delta x \approx 100 \times \Delta \log(x)$

Logaritmo natural

Elasticidad

Formas de medir la sensibilidad de y a las variaciones de x

- Pendiente de una línea recta depende de las unidades de medición de y y x
- Con la elasticidad, las variables no necesitan estar en las mismas unidades

La **elasticidad** mide el cambio porcentual en y cuando x aumenta 1%

$$\frac{\Delta y}{\Delta x} \frac{x}{y} = \frac{\% \Delta y}{\% \Delta x} \approx \frac{\Delta \log(y)}{\Delta \log(x)}$$

- Curva es inelástica si elasticidad < 1
- Curva es isoeástica si elasticidad $= 1$
- Curva es elástica si elasticidad > 1

Logaritmo natural

Elasticidad

La función log permite especificar *modelos con elasticidad constante*

$$\log(y) = \beta_0 + \beta_1 \log(x) \implies \beta_1 = \frac{\Delta \log(y)}{\Delta \log(x)}$$

- β_1 es la **elasticidad** de y con respecto a x
- Modelo ampliamente usado en economía empírica
- En $y = \beta_0 + \beta_1 x$, la elasticidad no es constante, depende de x

Pregunta

¿Cuál es la elasticidad $(\frac{\Delta y}{\Delta x} \frac{x}{y})$ de una línea recta?

Función exponencial

La función exponencial:

$$y = \exp(x) = e^x$$

La función exponencial es la inversa de la función logaritmo natural

- $\log(\exp(x)) = x, \quad \forall x$
- $\exp(\log(x)) = x, \quad x > 0$
- $\log(y) = \beta_0 + \beta_1 x$ es equivalente a $y = \exp(\beta_0 + \beta_1 x)$

Repaso de Probabilidad

Variables aleatorias

Experimento: Procedimiento que se puede repetir y que tiene un conjunto de resultados aleatorios bien definidos

- Ej. Lanzar una moneda n veces y contar el número de caras

Variable aleatoria: Valor numérico asignado al resultado de un experimento

- Permite el cálculo de probabilidades
- Se denotan con mayúsculas y el resultado particular con minúsculas ($X \neq x$)

Variables aleatorias discretas

X es discreta si toma valores numéricos finitos (o infinitamente contables) y cada valor está asociado con una probabilidad: $p_j = \mathbb{P}(X = x_j), j = 1, 2, \dots, k$

Ejemplo

La variable aleatoria Bernoulli (o binaria) toma los valores 1 (éxito) o 0 (fracaso)

- $X \sim \text{Bernoulli}(p), \mathbb{P}(X = 1) = p$ y $\mathbb{P}(X = 0) = 1 - p$

La **función de masa de probabilidad** (fmp) resume la información de una variable aleatoria discreta: $f_X(x_j) = p_j, j = 1, 2, \dots, k$

Propiedades de la fmp:

- Valores están entre 0 y 1 para todos los posibles eventos
- La suma de todos los valores de la fmp siempre es igual a 1 ($\sum_{i=1}^k p_j = 1$)

Variables aleatorias continuas

X es continua si toma cualquier valor en los números reales con probabilidad 0

La **función de densidad de probabilidad** (fdp) proporciona información sobre los valores que toma una variable aleatoria continua:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$$

- Área bajo la fdp entre los puntos a y b ($a < b$)
- Probabilidad de que X caiga entre a y b

Propiedades de la fdp:

- Valores están entre 0 y 1 para todos los posibles eventos
- El área total bajo la fdp siempre es igual a 1 ($\int f_X(x) dx = 1$)

Variables aleatorias continuas

fda

La **función de distribución acumulada** (fda) de una variable aleatoria X es:

$$F(X) = \mathbb{P}(X \leq x)$$

- Como es una probabilidad, siempre está entre 0 y 1
- Es una función de X no decreciente

Propiedades de la fda:

- Para cualquier número c , $\mathbb{P}(X > c) = 1 - F(c)$
- Para cualesquier números $a < b$, $\mathbb{P}(a \leq X \leq b) = F(b) - F(a)$

Vectores aleatorios

En econometría, nos interesa cómo una variable aleatoria Y se relaciona con otras variables aleatorias

- Cómo se comporta Y cuando las otras están sujetas a una condición
 - Ej. El valor de otra variable aleatoria es conocido

Un **vector aleatorio** se puede ver como un conjunto de variables aleatorias

- La distribución del vector aleatorio se denomina **distribución conjunta**
- A la distribución de cada componente se le denomina **distribución marginal**

Distribuciones conjuntas

La distribución conjunta de 2 variables aleatorias discretas X y Y está descrita por la **fdp conjunta** de (X, Y) : $f_{XY}(x, y)$

- Caso discreto: $f_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$

Las variables aleatorias X y Y son **independientes** si y solo si

$$f_{XY}(x, y) = f_X(x) f_Y(y), \quad \forall x, y,$$

de lo contrario son dependientes

- Independencia: Conocer el valor de una, no afecta la probabilidad de la otra
- $f_X(x)$ y $f_Y(y)$ se conocen como las **fdp marginales** de X y Y
- Caso discreto: $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \mathbb{P}(Y = y)$

Distribuciones condicionales

La **fdp condicional** resume la distribución condicional de Y dado X :

$$f_{Y|X}(y | x) = \frac{f_{XY}(x, y)}{f_X(x)}, \quad \forall x, f_X(x) > 0$$

Si X y Y son independientes,

$$f_{Y|X}(y | x) = f_Y(y)$$

y

$$f_{X|Y}(x | y) = f_X(x)$$

Características de las funciones de distribución

Utilizamos números o funciones (momentos) para capturar ciertas características

- Medidas de tendencia central: Formas de medir el centro de la distribución
 - Esperanza o valor esperado, mediana
- Medidas de dispersión: Miden qué tan lejos del centro están los valores
 - Varianza, desviación estándar
- Medidas de asociación: Miden la dependencia entre dos variables aleatorias
 - Covarianza, correlación, esperanza condicional

En econometría, utilizamos las propiedades de estos momentos, en lugar de calcularlos para distribuciones particulares (como en probabilidad)

Valor esperado o media poblacional (\mathbb{E})

Promedio ponderado de todos los valores de la variable aleatoria X

$$\mu_X = \mu = \mathbb{E}(X) = \sum_{j=1}^k x_j f_X(x_j) \quad \circ \quad \mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Propiedades (usadas en econometría):

- Para cualquier constante c , $\mathbb{E}(c) = c$
- Para cualesquier constantes a y b , $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- Para constantes $\{a_1, a_2, \dots, a_n\}$, $\mathbb{E}(\sum_{i=1}^n a_i x_i) = \sum_{i=1}^n a_i \mathbb{E}(x_i)$
 - Si $a_i = 1$, $\forall i$, la esperanza de la suma es la suma de las esperanzas

Preguntas

¿Cuál es el valor esperado de lanzar una moneda? ¿Y el de un dado?

Mediana (med)

El cálculo de la mediana depende del tipo de variable aleatoria

- Discreta: Ordenar de menor a mayor y escoger el valor medio
- Continua: Mitad del área a la izquierda y mitad a la derecha

Ambas $\mu_X = \mathbb{E}(X)$ y $med(X)$ son medidas válidas del centro de la distribución

- $\mu_X \neq med(X)$ en general
- $\mu_X = med(X)$ si la distribución es simétrica alrededor del valor μ

$$f(\mu + x) = f(\mu - x), \forall x$$

Varianza (Var)

Mide la distancia esperada de X a su media:

$$\sigma_X^2 = \sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}(X^2) - \mu^2$$

Propiedades:

- La varianza es no negativa
- Para cualquier constante c , $\text{Var}(c) = 0$
- Para cualesquier constantes a y b ,
 - $\text{Var}(aX + b) = a^2 \text{Var}(X)$
Sumar una constante, no altera la varianza
Multiplicar por una constante, incrementa la varianza
 - $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$
 - Si $\text{Cov}(X, Y) = 0$, $\text{Var}(X + Y) = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$

Desviación estándar (desvest)

$$\sigma_X = \sigma = \text{desvest}(X) = +\sqrt{\text{Var}(X)}$$

Propiedades:

- Para cualquier constante c , $\text{desvest}(c) = 0$
- Para cualesquier constantes a y b , $\text{desvest}(aX + b) = |a|\text{desvest}(X)$
- desvest está en las mismas unidades de medición que X (a diferencia de Var)

Pregunta

Si el salario promedio en miles de pesos es 52.3 y su desvest es 14.6, ¿cuál es el salario medio y su desvest en pesos?

Variable aleatoria **estandarizada**: $Z = (X - \mu)/\sigma$

Covarianza poblacional (Cov)

Mide la dependencia *lineal* entre 2 variables aleatorias

$$\sigma_{XY} = \text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}(XY) - \mu_X\mu_Y$$

- $\mu_X = \mathbb{E}(X)$, $\mu_Y = \mathbb{E}(Y)$
- Si $\sigma_{XY} > 0$, en promedio $X > \mu_X$ y $Y > \mu_Y$, o $X < \mu_X$ y $Y < \mu_Y$

Propiedades:

- Si X y Y son independientes, entonces $\text{Cov}(X, Y) = 0$
 - Pero $\text{Cov}(X, Y) = 0 \not\Rightarrow X$ y Y son independientes (ej. $Y = X^2$)
 - Entonces, correlación cero e independencia no son lo mismo
- Para cualesquier constantes a_1, a_2, b_1 y b_2 ,
 $\text{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\text{Cov}(X, Y)$
- Depende de las unidades de medida de las variables aleatorias

Correlación poblacional (Corr)

Mide la dependencia *lineal* pero no depende de las unidades de medida

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{desvest}(X) \text{desvest}(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Propiedades:

- $-1 \leq \rho_{XY} \leq 1, \forall X, Y$
- Si $X = Y$, $\text{Cov}(X, Y) = \text{Var}(X) = \sigma_X^2$ y $\rho_{XY} = 1$
- Si $\rho_{XY} = 0$, X y Y no correlacionadas (no hay relación lineal entre X y Y)
- Si $\rho_{XY} \approx 1$, la relación lineal es fuerte
- Para cualesquier constantes a_1, a_2, b_1 y b_2 ,
 - Si $a_1 a_2 > 0$, $\text{Corr}(a_1 X + b_1, a_2 Y + b_2) = \text{Corr}(X, Y)$
 - Si $a_1 a_2 < 0$, $\text{Corr}(a_1 X + b_1, a_2 Y + b_2) = -\text{Corr}(X, Y)$

Esperanza (o media) condicional

Una variable Y puede estar relacionada a una variable X de forma no lineal

La fdp condicional captura esa relación pero no se puede resumir en un solo valor numérico porque la distribución de Y dado $X = x$ depende de x

La **esperanza condicional** captura la relación no lineal entre Y y X

- Caso discreto: $\mathbb{E}(Y | X = x) = \sum_{j=1}^m y_j f_{Y|X}(y_j | X)$
- Promedio ponderado de posibles valores de Y dado que $X = x$
- Si cambia x , cambia $\mathbb{E}(Y | X)$ porque es una función de x

En econometría, capturamos la relación con funciones simples

- $\mathbb{E}(Y | X) = \alpha + \beta x$
- $\mathbb{E}(Y | X) = 10/x$

Esperanza condicional

Propiedades:

- $\mathbb{E}[c(X) | X] = c(X), \forall c(X)$
 - Funciones de X son como constantes
- Para funciones $a(X)$ y $b(X)$,
 $\mathbb{E}[a(X)Y + b(X) | X] = a(X)\mathbb{E}(Y | X) + b(X)$
- Si X y Y son independientes, $\mathbb{E}(Y | X) = \mathbb{E}(Y) \implies \text{Cov}(X, Y) = 0 (\neq)$
 - No depende de X
- Ley de las esperanzas iteradas: $\mathbb{E}[\mathbb{E}(Y | X)] = \mathbb{E}(Y)$
 - Si $\mathbb{E}[\mathbb{E}(U | X)] = \mathbb{E}(U)$ y $\mathbb{E}(U | X) = 0, \mathbb{E}(U) = 0$, entonces

$$\mathbb{E}(U | X) = \mathbb{E}(U) \implies \text{Cov}(X, U) = 0$$

Modelos probabilísticos

Distribuciones de probabilidad que dependen de parámetros

- Diferentes valores de los parámetros, generan diferentes distribuciones

Distribuciones comúnmente usadas en econometría:

- Distribución normal y normal estándar
- Distribución χ^2
- Distribución t
- Distribución F

Visualizador de distribuciones: [Aquí](#)

Distribución normal (N)

$$X \sim N(\mu, \sigma^2), \quad \text{donde} \quad \mu = \mathbb{E}(X) \quad \text{y} \quad \sigma^2 = \text{Var}(X)$$

- Distribución simétrica alrededor de μ
- Si $Y = \log(X)$ tiene una distribución N, X tiene una distribución **lognormal**
- $Z \sim N(0, 1)$ tiene una distribución **normal estándar**
 - $\phi(z)$ denota la fdp y $\Phi(z)$ denota la fda, entonces $\Phi(z) = \mathbb{P}(Z \leq z)$

Podemos usar $\Phi(z)$ para calcular la probabilidad de cualquier evento para Z

- $\mathbb{P}(Z > z) = 1 - \Phi(z)$
- $\mathbb{P}(Z < -z) = \mathbb{P}(Z > z)$
- $\mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a)$
- $\mathbb{P}(|Z| > c) = \mathbb{P}(Z > c) + \mathbb{P}(Z < -c) = 2[1 - \Phi(c)] \quad \forall c > 0$

Distribución N

Propiedades:

- Si $X \sim N(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim N(0, 1)$
- Si $X \sim N(\mu, \sigma^2)$, $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- Si X y Y tienen una distribución conjunta normal, entonces son independientes $\iff \text{Cov}(X, Y) = 0$
- Cualquier combinación lineal de variables aleatorias independientes con la misma distribución normal, tiene una distribución normal
- Si $Y_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$, $\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Ejercicio

Encuentra las probabilidades $\mathbb{P}(X \leq 6)$ y $\mathbb{P}(X > 4)$ si $X \sim N(5, 4)$

Distribución χ^2

Si $Z_i \stackrel{\text{iid}}{\sim} N(0, 1)$, $i = 1, 2, \dots, n$, y $X = \sum_{i=1}^n Z_i^2$, entonces

$$X \sim \chi_n^2$$

- Distribución no negativa y no simétrica
- Su forma cambia con n
- n son los grados de libertad
- Se utiliza en la construcción de las distribuciones t y F

Distribución t

Si $Z \sim N(0, 1)$ y $X \sim \chi_n^2$ son independientes, entonces

$$T = \frac{Z}{\sqrt{X/n}} \sim t_n$$

- Forma similar a la distribución normal estándar, pero más dispersa
 - Valores extremos de T son más probables que los de Z
 - La diferencia entre ambas distribuciones disminuye cuando $n \rightarrow \infty$ porque $t_n \rightarrow Z$
- n son los grados de libertad
- Se utiliza para pruebas de hipótesis de un parámetro

Distribución F

Si $X_1 \sim \chi_{k_1}^2$ y $X_2 \sim \chi_{k_2}^2$ son independientes, entonces

$$F = \frac{X_1/k_1}{X_2/k_2} \sim F_{k_1, k_2}$$

- Distribución no negativa
- k_1 y k_2 son los grados de libertad
 - Orden de k_1 y k_2 es importante
- Se utiliza para hacer pruebas de hipótesis de varios parámetros
 - k_1 es el número de restricciones en los parámetros
 - k_2 es el número de observaciones menos los parámetros estimados
- $t_n^2 \sim F_{1, n}$

Repaso de Estadística

Inferencia estadística

La estadística inferencial estudia la estimación (puntual, por intervalos) de parámetros y pruebas de hipótesis sobre esos parámetros

- Requiere identificar la población de interés y especificar el modelo (fdp)

Ejemplo

Variable aleatoria Y con fdp conocida, $f(y; \theta)$, pero parámetro θ desconocido

- Diferentes valores de θ dan lugar a diferentes distribuciones de probabilidad

Pregunta

¿Cuáles son los parámetros en $Y_i \sim \text{Bernoulli}(p)$ y $Y_i \sim N(\mu, \sigma^2)$?

Muestra aleatoria

Si Y_1, Y_2, \dots, Y_n son variables aleatorias independientes con la misma fdp $f(y; \theta)$, entonces $\{Y_1, Y_2, \dots, Y_n\}$ es una **muestra aleatoria** de la población representada por $f(y; \theta)$ y se denota como i.i.d.

- Datos observados cambian con cada muestra

Tipo de muestreo común en estadística

- Estimadores puntuales (propiedades)
 - Muestras finitas: falta de sesgo, eficiencia
 - Muestras grandes: consistencia, distribución asintótica
- Estimación por intervalos
- Pruebas de hipótesis

Estimadores y estimados

Dada $\{Y_1, Y_2, \dots, Y_n\}$ de una fdp con parámetro θ (desconocido), un estimador W de θ es una regla que asigna a cada muestra particular un valor para θ

$$W = \hat{\theta} = h(Y_1, Y_2, \dots, Y_n) \quad \text{vs} \quad w = \hat{\theta} = h(y_1, y_2, \dots, y_n)$$

- W es una v.a. (tiene fdp, $\mathbb{E}(\cdot)$, $\text{Var}(\cdot)$) y se llama **estimador** puntual
- w es un número y se llama **estimado** puntual

Ejemplo

Dada $\{Y_1, Y_2, \dots, Y_n\}$ de una población con media μ y varianza σ^2 , el *estimador* de μ es $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ y el de σ^2 es $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$. Para una muestra particular $\{y_1, y_2, \dots, y_n\}$, el *estimado* de \bar{Y} es $\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$ y el de S^2 es $s^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n-1}$

Distribución muestral

La distribución de probabilidad del estimador W se llama **distribución muestral**

- Objeto de estudio de la estadística matemática
- Énfasis en características de la distribución muestral de W para evaluarlo como estimador de θ

La distribución muestral depende de la distribución de Y y de la función h

- La distribución de Y está fuera de nuestro control
- Pero podemos escoger la función h (supuestos, método de estimación)

Hay muchas formas de generar estimadores

- Queremos propiedades que nos permitan descartar los irrelevantes

Falta de sesgo (insesgadez)

El sesgo de un estimador W de θ se define como

$$\text{Sesgo}(W) = \mathbb{E}(W) - \theta$$

- W es un **estimador insesgado** de θ si $\mathbb{E}(W) = \theta$, $\forall \theta$
 - Promedio de infinitas muestras da θ , aunque en la práctica una muestra

Ejemplo

$$\mathbb{E}(\bar{Y}) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{1}{n} n \mu = \mu \quad \text{y} \quad \mathbb{E}(S^2) = \sigma^2$$

Pregunta

Si $W = Y_1$, entonces $\mathbb{E}(Y_1) = \mu$, ¿es útil?

Varianza muestral

Podemos tener 2 estimadores insesgados, pero uno con menor dispersión

- ¿Cuál preferimos?

La **varianza muestral** es la varianza de un estimador $\text{Var}(W)$

- Número (desconocido) que mide dispersión de la distribución muestral de W

Ejemplo

$$\text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} \rightarrow 0 \text{ cuando } n \rightarrow \infty$$

Pregunta

Si $W = Y_1$, ¿ $\text{Var}(Y_1)$? ¿Se prefiere $W = \bar{Y}$ o $W = Y_1$ para estimar μ ?

Eficiencia relativa

Si W_1 y W_2 son estimadores insesgados de θ , W_1 es **eficiente** relativo a W_2 si

- $\text{Var}(W_1) \leq \text{Var}(W_2) \quad \forall \theta$, y
- $\text{Var}(W_1) < \text{Var}(W_2)$ para al menos un valor de θ

Restringimos nuestra atención a estimadores insesgados generalmente

Pero para comparar estimadores sesgados usamos el **error cuadrático medio**

$$\mathbb{E}[(W - \theta)^2] = \text{Var}(W) + [\text{Sesgo}(W)]^2$$

- Ej. Tirador de flechas a distancia (sesgo vs eficiencia)

Consistencia

Razonable requerir que el procedimiento de estimación mejore conforme $n \rightarrow \infty$

- Análisis asintótico aproxima la distribución muestral de W con base en n

Si W_n es un estimador de θ obtenido de una muestra aleatoria Y_1, Y_2, \dots, Y_n ,

W_n es un estimador **consistente** de θ si $\forall \varepsilon > 0$,

$$\mathbb{P}(|W_n - \theta| > \varepsilon) \rightarrow 0 \quad \text{cuando} \quad n \rightarrow \infty$$

- Se dice que θ es el límite en probabilidad de W_n : $\text{plim}(W_n) = \theta$
- La distribución de W_n se concentra cada vez más alrededor de θ
- Consistencia es un requerimiento mínimo de un estimador

Si W_n es un estimador insesgado de θ y $\text{Var}(W_n) \rightarrow 0$ cuando $n \rightarrow \infty$, entonces $\text{plim}(W_n) = \theta$; es decir, W_n es consistente

Ley de los Grandes Números (LGN)

Sean Y_1, Y_2, \dots, Y_n variables aleatorias i.i.d con media μ , entonces

$$\text{plim} (\bar{Y}_n) = \mu$$

- \bar{Y}_n es consistente para μ
- Podemos acercarnos a μ con una muestra suficientemente grande

Se puede combinar con las propiedades de los plim para mostrar que diferentes estimadores (funciones de otros) son consistentes

Ejemplo

$S_n = \sqrt{S_n^2}$ desviación estándar muestral no es insesgado pero es consistente

Normalidad asintótica

Necesitamos la forma de la distribución muestral de W_n para

- Construir intervalos de confianza
- Hacer pruebas de hipótesis

Si $\{Z_n \mid n = 1, 2, \dots\}$ es una secuencia de variables aleatorias tal que $\forall z$, $\mathbb{P}(Z_n \leq z) \rightarrow \Phi(z)$ cuando $n \rightarrow \infty$, entonces

$$Z_n \overset{a}{\sim} N(0, 1)$$

- Se dice que Z_n tiene una **distribución normal estándar asintótica**

En muestras grandes, la distribución muestral de muchos estimadores es aproximadamente normal

Teorema del Límite Central (TLC)

Sea $\{Y_1, Y_2, \dots, Y_n\}$ una muestra aleatoria con media μ y varianza σ^2 , entonces

$$Z_n = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0, 1)$$

- El promedio de una m.a. de *cualquier* población (con varianza finita), cuando se estandariza, tiene una distribución normal estándar asintótica
- Muchos estimadores se pueden escribir como funciones de promedios muestrales, por lo que podemos aplicar la LGN y el TLC

Ejemplo

$\frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}}$ es un estadístico que es asintóticamente normal

De aquí en adelante, ya no se incluye el subíndice n

Métodos de estimación puntual

Métodos de estimación de parámetros comúnmente usados:

- Método de momentos
 - Momento poblacional en función de parámetro se iguala al muestral
- Máxima verosimilitud
 - Máximo de una función tal que los datos observados son los más probables
- Mínimos cuadrados
 - Encuentran la línea que ajusta mejor la dirección general de los datos

Estos métodos producen estimadores con buenas propiedades

- Insesgados, eficientes, consistentes

Estimación por intervalos (de la media poblacional)

¿Qué tan cerca está un estimado del parámetro θ ?

- No sabemos porque no conocemos θ , pero podemos usar probabilidades
- La desviación estándar muestral evalúa la incertidumbre de un estimador

Sea $\{Y_1, Y_2, \dots, Y_n\}$ una muestra aleatoria de $N(\mu, \sigma^2)$. Si usamos \bar{Y} y S^2 ,

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Si c es el percentil 97.5 de la distribución t_{n-1} tal que $\mathbb{P}(-c < t_{n-1} < c) = 0.95$

$$\mathbb{P}\left(-c < \frac{\bar{Y} - \mu}{S/\sqrt{n}} < c\right) = \mathbb{P}\left(\bar{Y} - c\frac{S}{\sqrt{n}} < \mu < \bar{Y} + c\frac{S}{\sqrt{n}}\right) = 0.95$$

Estimación por intervalos

El siguiente **intervalo de confianza** (IC) contiene a μ 95% de las veces

$$\left[\bar{Y} - c \frac{S}{\sqrt{n}}, \bar{Y} + c \frac{S}{\sqrt{n}} \right]$$

- Para una muestra, el IC es $\left[\bar{y} - c \frac{s}{\sqrt{n}}, \bar{y} + c \frac{s}{\sqrt{n}} \right]$

Si c_α denota el percentil $100(1 - \alpha)$ de la distribución t_{n-1} , entonces el IC para una muestra particular es $\left[\bar{y} - c_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{y} + c_{\alpha/2} \frac{s}{\sqrt{n}} \right]$

- c_α requiere escoger α y conocer los grados de libertad $n - 1$
- c_α se puede obtener de las tablas de la distribución

Estimación por intervalos

Dado que $\text{desvest}(\bar{y}) = \frac{\sigma}{\sqrt{n}}$, la variable aleatoria $\frac{S}{\sqrt{n}}$ es el **error estándar** de \bar{Y}

- Si definimos el error estándar del estimado puntual de \bar{y} como $\text{errest}(\bar{y}) = \frac{s}{\sqrt{n}}$, podemos reescribir el IC como $[\bar{y} \pm c_{\alpha/2} \times \text{errest}(\bar{y})]$
- Dado que $t_n \rightarrow N(0, 1)$, para $\alpha = 0.05$, $c_{\alpha/2} \rightarrow 1.96$ cuando $n \rightarrow \infty$, entonces la regla general para un IC al 95% es $[\bar{y} \pm 2 \times \text{errest}(\bar{y})]$
- Cuando la población tenga una distribución no normal, pero un tamaño de muestra grande, podemos usar el TLC para aproximar la distribución de \bar{Y}
 - Un IC al 95% aproximado es $[\bar{y} \pm 1.96 \times \text{errest}(\bar{y})]$

Dispersión en distribución poblacional y muestral

- σ es la desviación estándar poblacional (desconocida) de Y
- $s = \hat{\sigma}$ es el estimado de σ
- $\text{desvest}(\bar{Y}) = \frac{\sigma}{\sqrt{n}}$ es la desviación estándar muestral de \bar{Y}
 - $\text{desvest}(\bar{Y}) \rightarrow 0$ cuando $n \rightarrow \infty$
 - El estimador \bar{Y} de μ se vuelve más preciso conforme $n \rightarrow \infty$
- $\widehat{\text{desvest}}(\bar{Y}) = \text{errest}(\bar{Y}) = \frac{s}{\sqrt{n}}$ es el error estándar de \bar{Y}
 - errest es un estimado de $\frac{\sigma}{\sqrt{n}}$
 - Es lo que se usa para construir IC
 - Es fundamental para pruebas de hipótesis

Pruebas de hipótesis

Métodos para contestar preguntas binarias utilizando una muestra de datos

- ¿Qué tan fuerte es la evidencia en contra de una hipótesis?

Tipos de hipótesis:

- La **hipótesis nula** H_0 se asume cierta hasta que los datos sugieran que hay suficiente evidencia en favor de la **hipótesis alternativa** H_1 o H_a
- Hay hipótesis alternativas *de un lado* (evidencia contra H_0 en una sola dirección) o *de dos lados* (cualquier desviación de H_0)

Tipos de errores:

- Error **tipo I**. Rechazar H_0 cuando es en realidad cierta
- Error **tipo II**. No rechazar H_0 cuando es en realidad falsa

Pruebas de hipótesis

Al decidir sobre H_0 , podemos acertar o equivocarnos y nunca sabremos cuál fue

- Pero podemos calcular la probabilidad de cometer cualquiera de los 2 errores

Construimos las pruebas tal que la probabilidad de cometer error tipo I sea baja

- El **nivel de significancia** de la prueba α es la probabilidad del error tipo I
 - $\alpha = \mathbb{P}(\text{Rechazar } H_0 \mid H_0)$
 - Valores comunes para α son $\{0.1, 0.05, 0.01\}$
 - Cuantifica la tolerancia de cometer error tipo I
- Minimizamos el error tipo II al maximizar el **poder de la prueba**
 - $\pi(\theta) = \mathbb{P}(\text{Rechazar } H_0 \mid \theta) = 1 - \mathbb{P}(\text{Tipo II} \mid \theta)$

Pruebas de hipótesis

Para probar H_0 contra H_a , se necesita un estadístico de prueba y un valor crítico

- Un **estadístico de prueba** T es una variable aleatoria que asigna un valor a cada muestra
 - Cuantifica cuánto se desvían los datos observados de lo esperado bajo H_0
 - Diferencia con estimador: Un estimador no depende de θ
 - Para una muestra particular es t
- Tras definir el nivel de significancia α , encontramos el **valor crítico** c con base en la distribución de T bajo el supuesto de que H_0 es cierta
- La **regla de rechazo** determina cuándo H_0 se rechaza al comparar el estadístico de prueba t contra un valor crítico c
 - Lenguaje: No se rechaza H_0 (✓) vs se acepta H_0 (✗)

Prueba de hipótesis de la media de una población normal

Prueba de un lado para μ de una población $N(\mu, \sigma^2)$

$$H_0 : \mu = \mu_0 \text{ (ej. } \mu_0 = 0)$$

$$H_a : \mu > \mu_0$$

Estadístico t

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \implies t = \frac{\bar{y} - \mu_0}{\text{errest}(\bar{y})}$$

Si $\alpha = 0.05$, c se escoge tal que la probabilidad de error tipo I es 5%:

$$\mathbb{P}(T > c \mid H_0) = 0.05$$

Regla de rechazo: $t > c$, donde c es el percentil $100(1 - \alpha)$ de distribución t_{n-1}

Prueba de hipótesis de la media de una población normal

H_a	Escoger c tal que	Regla rechazo	c percentil	valor $-p$
$\mu > \mu_0$	$\mathbb{P}(T > c \mid H_0) = \alpha$	$t > c$	$100(1 - \alpha)$	$\mathbb{P}(T > t)$
$\mu < \mu_0$	$\mathbb{P}(T < -c \mid H_0) = \alpha$	$t < -c$	$100(1 - \alpha)$	$\mathbb{P}(T < t)$
$\mu \neq \mu_0$	$\mathbb{P}(T > c \mid H_0) = \alpha$	$ t > c$	$100(1 - \frac{\alpha}{2})$	$\mathbb{P}(T > t)$

Prueba de hipótesis de la media de una población normal

Cuando la población no es normal pero $n \rightarrow \infty$, por el TLC,

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \underset{a}{\sim} N(0, 1)$$

- t se calcula como antes pero se utilizan valores críticos de $N(0, 1)$

Los IC se pueden usar para hacer PH de dos lados

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0$$

- Si $\mu_0 \notin IC$, H_0 se rechaza (depende del valor de α)

Valores p

El **valor p** indica el nivel de significancia más alto para el que no se rechaza H_0

- Asociado con el valor del estadístico como nuestro valor crítico
- Probabilidad de valores más extremos al valor del estadístico si H_0 es cierta

Ejemplo

Si $H_0 : \mu = \mu_0$, $H_a : \mu > \mu_0$ y $T \overset{a}{\sim} N(0, 1)$,

$$\text{valor} - p = \mathbb{P}(T > t \mid H_0) = 1 - \Phi(t) = 0.065$$

- Si $\alpha < 0.065$, H_0 no se rechaza . Si $\alpha > 0.065$, H_0 se rechaza

Valores p pequeños generalmente son evidencia en contra de H_0

- Regla de rechazo: $\text{valor} - p < \alpha$
- Permite hacer una prueba de hipótesis para cualquier nivel de significancia

Comentarios

Es importante considerar significancia tanto práctica como estadística

- Los estimados puntuales indican dirección (signo) y magnitud

Para entender conceptos en inferencia, es importante distinguir entre

- Estimador W (variable aleatoria) y estimado w (número)

En **econometría**, la convención es utilizar

- θ para un parámetro poblacional, y
- $\hat{\theta}$ tanto para el estimador como para el estimado

El contexto indica si se trata del estimador (propiedades estadísticas) o del estimado (número)