Índice general

Lección 1. Introducción a la econometría

Marcos Bujosa

19 de septiembre de 2026

Resumen

Breve introducción: ¿qué es la econometría? ¿para qué sirve? algunos ejemplos, tipología de datos y enfoque del curso.

"La econometría es la medida de lo económico." — Ragnar Frisch

- (slides) — (html) — (pdf)

1 ¿Qué es la econometría?

- Disciplina que *mide* relaciones económicas con datos.
  - Convierte afirmaciones cualitativas en cuantitativas:

    "a una mayor educación corresponden salarios más altos" $\longrightarrow$ "un año más de educación $\approx 8\%$ más de salario".
- Confluencia de tres disciplinas:
  - Teoría económica (qué relaciones son relevantes).
  - Estadística (cuantifica el margen de error de nuestras estimaciones).
  - Matemáticas / geometría de $\mathbb{R}^n$ (lenguaje formal).
- Cuarto pilar práctico: software (en este curso, *Gretl*).

¿Qué es la econometría?

Esta asignatura será, para muchos, su primer contacto con la econometría. El término puede resultar intimidante, pero conviene desactivar la inquietud cuanto antes: la econometría no es una disciplina cerrada, ni un conjunto de recetas matemáticas oscuras, ni mucho menos un terreno reservado a quienes tengan especial habilidad con las fórmulas. Es, simplemente, el conjunto de técnicas que permiten *medir* relaciones económicas a partir de datos reales y juzgar si la evidencia respalda, contradice o matiza las afirmaciones que las teorías económicas hacen sobre el mundo.

Ragnar Frisch, economista noruego y primer Premio Nobel de Economía (compartido con Jan Tinbergen en 1969), acuñó el término *econometría* en 1926 al fundar la *Econometric Society*. La

cita que aparece en la portada — "la econometría es la medida de lo económico" — resume la vocación de la disciplina: cuantificar lo que la teoría afirma en términos cualitativos. Convertir "más educación, más salario" en una cifra concreta que diga *cuánto* más, en qué condiciones y con qué grado de incertidumbre.

Este curso está pensado para un perfil heterogéneo: estudiantes de PPE con bagaje variado en matemáticas. Por eso adoptaremos un enfoque que prioriza la *comprensión geométrica* sobre la manipulación algebraica. Verá que conceptos aparentemente abstractos (media, varianza, regresión) tienen interpretaciones visuales muy concretas, y que entender esas interpretaciones le ahorrará después memorizar fórmulas que sin contexto resultan opacas.

Para profundizar:

- Bjerkholt, O. (2017). On the founding of the Econometric Society. *Journal of the History of Economic Thought*, 39(2), 175–198. Historia accesible del nacimiento de la disciplina.
- Vídeo corto: "Mastering Econometrics with Joshua Angrist — Marginal Revolution University. Joshua Angrist, premio Nobel 2021.

2 ¿Para qué sirve la econometría?

Tres usos, conceptualmente distintos:

1. *Descripción*: resumir asociaciones en una muestra.
2. *Predicción*: anticipar valores no observados.
3. *Inferencia causal*: identificar relaciones de causa-efecto.

En este curso nos centramos en (1) y (2), y sentamos las bases de (3).

Advertencia central del curso:

Correlación no implica causalidad.

¿Para qué sirve la econometría?

Conviene distinguir con claridad tres usos de la econometría, porque se mezclan habitualmente en el discurso público pero son conceptualmente muy distintos y exigen condiciones diferentes.

El primer uso es la *descripción*. Aquí simplemente queremos resumir cómo se relacionan dos o más variables en una muestra. Si observamos a 1.000 trabajadores y constatamos que, en promedio, los que tienen un año más de educación ganan un 8% más, estamos describiendo una asociación observada en esa muestra. No estamos afirmando que la educación cause el salario, ni que en otra muestra observaríamos lo mismo. Solo estamos resumiendo lo que vemos. La descripción es útil y suele ser el primer paso de cualquier análisis.

El segundo uso es la *predicción*. Aquí queremos anticipar el valor de una variable a partir de los valores de otras. Por ejemplo, dado el historial crediticio de un cliente, ¿qué probabilidad tiene de dejar de pagar un préstamo? El banco no necesita saber *por qué* ciertos perfiles incurren con mayor frecuencia en el impago; le basta con saber que *predicen* bien el impago para tomar una decisión

informada. La predicción es la tarea central de buena parte del aprendizaje automático moderno y comparte muchas herramientas con la econometría.

El tercer uso, y el más ambicioso, es la *inferencia causal*. Aquí queremos determinar si una variable *causa* cambios en otra, no solo si están asociadas. Esta tarea es delicada porque la asociación puede tener orígenes muy distintos: causalidad directa, causalidad inversa, factores comunes que afectan a ambas variables, sesgos de selección, etc.

La frase *correlación no implica causalidad* es probablemente la advertencia metodológica más conocida en estadística. Un ejemplo clásico: en muchas ciudades, hay una correlación positiva entre el número de policías por habitante y la tasa de criminalidad. Esto no significa que los policías causen el crimen; significa que las ciudades con más crimen contratan más policías.

En este curso nos centraremos en los dos primeros usos. Tras el curso será capaz de describir relaciones entre variables y de predecir valores con cierta solvencia. Para hacer inferencia causal con rigor necesitará herramientas adicionales (variables instrumentales, diferencias en diferencias, regresión discontinua, experimentos aleatorios) que se ven en cursos posteriores, especialmente en *microeconometría* o *evaluación de políticas públicas*.

Para profundizar:

- Angrist, J. y Pischke, J.-S. (2014). *Mastering "Metrics": The Path from Cause to Effect*. Un libro corto, accesible y muy bien escrito sobre inferencia causal. Capítulo 1: ideal para entender la distinción asociación / causalidad.
- Vídeo: "Correlation is not Causation", de Spurious Correlations, basado en el libro de Tyler Vigen. Muestra correlaciones absurdas (consumo de queso y muertes por enredo con sábanas) para fijar la idea.
- Pearl, J. y Mackenzie, D. (2018). *The Book of Why*. Lectura divulgativa pero rigurosa sobre por qué la inferencia causal es difícil y cómo abordarla.

3 Ejemplos motivadores

- *Salarios*: ¿cuánto rinde un año más de educación? (ecuación de Mincer).
- *Demanda*: ¿cuánto cae la cantidad de tabaco demandada si sube el precio un 10%?
- *Crecimiento*: ¿por qué unos países crecen más? Instituciones, capital humano, apertura.
- *Políticas públicas*: ¿un programa de formación aumenta realmente el empleo?
- *Vivienda*: ¿qué características explican el precio? (lo veremos en Gretl con *hprice2*).

Ejemplos motivadores

A lo largo del curso volveremos sobre varios ejemplos que conviene tener presentes desde el principio. No es necesario que entienda ahora los detalles; basta con que se queden en la memoria como anclas concretas para los conceptos abstractos que iremos introduciendo.

La *ecuación de salarios* o *ecuación de Mincer* es el ejemplo canónico. Jacob Mincer propuso en 1974 una especificación sencilla en la que el logaritmo del salario depende linealmente de los años de educación y de un polinomio en la experiencia laboral. Esta ecuación, en sus variantes, se ha estimado en miles de estudios y proporciona la base para discutir los rendimientos de la educación, la discriminación salarial por género o raza, las primas por sector, etc. La veremos formalmente cuando estudiemos la regresión múltiple.

La *función de demanda* es otro caso clásico. La teoría microeconómica predice cómo varía la cantidad demandada de un bien ante cambios en su precio, en la renta y en los precios de bienes relacionados. La econometría permite estimar esas relaciones a partir de datos de mercado. El tabaco, la gasolina, los alimentos básicos y, más recientemente, los bienes digitales han sido objeto de estudios de demanda intensos, tanto por su interés académico como por sus implicaciones para la política fiscal (impuestos especiales) y la salud pública.

Los *determinantes del crecimiento económico* ocupan a la macroeconomía empírica desde hace décadas. Tras los trabajos seminales de Robert Barro y otros, se ha intentado identificar qué variables explican que unos países crezcan más que otros: el capital humano, las instituciones, la apertura comercial, la geografía, la corrupción. Es un campo lleno de controversias metodológicas porque la inferencia causal con datos de países es muy difícil: hay pocas observaciones, muchas variables candidatas y la causalidad puede ir en cualquier dirección (¿las buenas instituciones causan crecimiento, o el crecimiento permite construir buenas instituciones?).

La *evaluación de políticas públicas* es probablemente el ámbito donde la econometría moderna ha dado sus frutos más visibles. ¿Un programa público de formación para parados aumenta realmente sus posibilidades de encontrar empleo? La respuesta no es trivial: quienes se apuntan al programa pueden ser, de entrada, más motivados que quienes no lo hacen, y por tanto encontrarían empleo más fácilmente incluso sin el programa. Separar el efecto *del programa* del efecto *de la selección* es el problema central de la evaluación de impacto. Los premios Nobel de 2019 (Banerjee, Duflo, Kremer) y 2021 (Card, Angrist, Imbens) reconocen contribuciones en esta línea.

El último ejemplo, los *precios de la vivienda*, es el que veremos en la primera sesión de laboratorio. Usaremos el conjunto de datos *hprice2*, que viene preinstalado en Gretl y corresponde al célebre estudio de Harrison y Rubinfeld (1978) sobre precios de la vivienda en el área metropolitana de Boston. El objetivo del estudio original era estimar cuánto estaban dispuestos a pagar los residentes por una mejora en la calidad del aire (medida por la concentración de óxidos de nitrógeno). Es un ejemplo precioso porque combina econometría, valoración ambiental y política pública.

Para profundizar:

- Mincer, J. (1974). *Schooling, Experience and Earnings*. NBER. El libro fundacional.
- Heckman, J., Lochner, L. y Todd, P. (2006). *Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. Handbook of the Economics of Education*, vol. 1. Revisión moderna.
- Harrison, D. y Rubinfeld, D. L. (1978). *Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management*, 5(1), 81–102. El artículo que originó el dataset que usaremos.
- Acemoglu, D., Johnson, S. y Robinson, J. (2001). *The colonial origins of comparative deve-*

lopment. American Economic Review, 91(5), 1369–1401. Ejemplo célebre (y polémico) en la literatura del crecimiento.

- Duflo, E. (2017). *The economist as plumber. American Economic Review*, 107(5), 1–26. Charla accesible sobre el papel del economista empírico en el diseño de políticas.

4 La idea clave: los datos como vectores

Podemos escribir n mediciones de una variable como una *lista ordenada de n números*:

$$\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n.$$

- Cada dato y_i corresponde a una medición identificada por el índice i.
- Los tipos de datos dependen del *significado del índice* y la *relevancia del orden.*
- Una *lista ordenada de n números* se denomina vector de $\mathbb{R}^n$.
- Esta visión vectorial es la puerta de entrada al enfoque geométrico del curso.

La idea clave: los datos como vectores

Esta es la transparencia más importante de la sesión, y la que conviene leer con calma. Es la puerta de entrada al enfoque geométrico que vertebrará todo el curso.

Cuando observamos una variable económica — por ejemplo, el salario — en n individuos, lo que obtenemos es una *lista ordenada de n números.* Si llamamos $\boldsymbol{y}$ a esa lista, podemos escribirla como

$$\boldsymbol{y} = (y_1, y_2, \ldots, y_n),$$

donde y_i es el salario del individuo i. Cada componente y_i es un número real, y la lista completa tiene n posiciones. En lenguaje matemático, $\boldsymbol{y}$ es un *vector* del espacio $\mathbb{R}^n$.

La palabra "vector" puede evocar, si recuerda el bachillerato, una flecha (en el plano o en el espacio) que parte del origen de coordenadas y señala un punto del espacio. Esa imagen es útil para $n = 2$ o $n = 3$, donde literalmente podemos dibujar la flecha y situar fielmente el punto. Para n mayor, la imagen geométrica deja de ser visualizable, pero las operaciones matemáticas son *las mismas*: sumamos vectores componente a componente, los multiplicamos por escalares, calculamos productos escalares y normas con las mismas fórmulas. Lo único que perdemos es la posibilidad de dibujar fielmente; no perdemos ni la estructura ni la intuición, y seguiremos usando dibujos esquemáticos.

Esta es una de las ideas más liberadoras de la matemática moderna: el mismo *tipo* de objeto (una lista de números) admite el mismo *tipo* de operaciones independientemente de cuántas componentes tenga. Cuando en las próximas sesiones hablemos de "el ángulo entre dos vectores de $\mathbb{R}^{1000}$", no estaremos haciendo nada exótico: estaremos aplicando, a una lista de 1.000 números, la misma fórmula que define el ángulo entre dos flechas en el plano.

¿Por qué insistir en esta idea? Porque toda la econometría que veremos consistirá en hacer geometría con vectores de este tipo. La media de una variable será una *proyección* (la sombra del

vector de datos sobre una recta especial). La varianza saldrá del *teorema de Pitágoras*. La regresión lineal será también una proyección, sobre un plano o sobre un subespacio de mayor dimensión. El coeficiente de determinación R^2 será el cuadrado del coseno de un ángulo (o un ratio entre dos longitudes). Estas afirmaciones, ahora opacas, se irán haciendo transparentes a medida que avancemos.

Conviene retener un ejemplo numérico concreto. Supongamos cinco trabajadores y sus salarios mensuales en euros:

$$\boldsymbol{y} = (1800, 2200, 1500, 2800, 2000).$$

Esto es un vector de $\mathbb{R}^5$. Si añadiéramos su edad,

$$\boldsymbol{x} = (28, 45, 22, 52, 35),$$

tendríamos otro vector de $\mathbb{R}^5$. Toda la econometría del curso consistirá en preguntarse cosas como: ¿qué relación hay entre $\boldsymbol{y}$ y $\boldsymbol{x}$? ¿Cuál es la mejor aproximación lineal de $\boldsymbol{y}$ a partir de $\boldsymbol{x}$? ¿Cuán parecida es la *dirección* de $\boldsymbol{y}$ a la *dirección* de $\boldsymbol{x}$? Preguntas geométricas, todas, formuladas sobre vectores.

Para profundizar:

- Strang, G. (2016). *Introduction to Linear Algebra* (5ª ed.), capítulo 1. Aunque es un texto de álgebra lineal, los primeros capítulos están escritos con una intuición geométrica admirable.
- Vídeo: serie "Essence of Linear Algebra" de 3Blue1Brown. Especialmente recomendados los dos primeros capítulos. Las animaciones hacen ver lo que los libros solo describen.
- Wooldridge, J. M. (2020), apéndice D, secciones D.1 y D.2. Repaso de vectores en notación más austera.
- Vídeo: Curso de Álgebra Lineal. Primeros vídeos de mi curso de álgebra lineal. Marcos Bujosa

5 Datos de sección cruzada

- Índice i = identifica una persona, empresa, país, región.
- Normalmente todas las observaciones corresponden al mismo momento.
- El orden del índice es *arbitrario*: podríamos reindexar y nada cambiaría.
- Pero una vez fijado un orden, debe mantenerse en *todas* las variables:
  - si y_1 es el salario de Ana, x_1 debe ser la educación de Ana.

Ejemplo: salario, educación y edad de 1.000 trabajadores en 2024.

Datos de sección cruzada

Los *datos de sección cruzada* (en inglés, *cross-sectional data*) son aquellos en los que cada observación corresponde a una unidad distinta — una persona, una empresa, un país, una región—, y normalmente las observaciones se refieren al mismo momento temporal (y, en cualquier caso, el orden cronológico de los datos no juega ningún papel).

El rasgo característico de este tipo de datos es que *el orden de las observaciones en el vector es arbitrario*. Si encuestamos a 1.000 trabajadores, la asignación de los índices del 1 al 1.000 es indiferente: podríamos numerarlos según el orden de llegada, por sorteo, alfabéticamente o por su documento de identidad. Cambiar el orden de la lista según cualquiera de estos criterios no altera en absoluto el análisis. El índice i identifica unívocamente a una persona, pero el orden en que las personas aparecen en el conjunto de datos no arroja información relevante en el análisis.

Eso sí, una vez que hemos fijado un orden, debemos *mantenerlo consistentemente entre todas las variables*. Si en la posición $i = 1$ está Ana, entonces en la primera componente del vector de salarios debe estar el salario de Ana, en la primera componente del vector de educación deben estar los años de educación de Ana, en la primera componente del vector de edad debe estar la edad de Ana, y así sucesivamente. Romper esta consistencia es uno de los errores más graves (y, lamentablemente, más frecuentes) en el manejo de datos: si reordena los datos de una variable y olvida reordenar las otras del mismo modo, todo el análisis pierde sentido. Gretl y los demás programas estadísticos manejan los datos por filas precisamente para evitar este problema, pero conviene tenerlo presente.

Esta arbitrariedad del orden es compatible con el enfoque geométrico del curso. Todas las operaciones que vamos a hacer con vectores — sumar componente a componente, calcular productos escalares, calcular normas — arrojan el *mismo resultado* si reordenamos las componentes (siempre que reordenemos todos los vectores de la misma manera). En lenguaje técnico: nuestras operaciones son invariantes ante permutaciones de los índices. Por eso $\mathbb{R}^n$ es el "hogar natural" para datos de sección cruzada.

Los ejemplos más habituales son encuestas a hogares o individuos (la Encuesta de Población Activa en España, la European Social Survey), datos de empresas en un año concreto (Central de Balances del Banco de España) o datos de países en un momento dado (indicadores del Banco Mundial para un año). El conjunto *hprice2* de Gretl que usaremos en la primera sesión de laboratorio es un ejemplo perfecto: 506 áreas censales del Boston metropolitano observadas en torno a 1970, sin componente temporal.

Para profundizar:

- Wooldridge, J. M. (2020), capítulo 1, sección 1.3. Discusión clara de los tipos de datos.
- INE: la Encuesta de Población Activa y la Encuesta de Condiciones de Vida. Ejemplos vivos de datos de sección cruzada en España.
- Banco Mundial: World Development Indicators. Permite descargar datos de países para cualquier año.

6 Series temporales

- Índice t = momento del tiempo (enero 2020, febrero 2020, …).
- El orden *sí* contiene información: febrero va después de enero.
- No tiene sentido reordenar.

Ejemplo: tasa de paro mensual en España, 2000–2024.

- En este curso usaremos series temporales *como si fueran* sección cruzada.

- Esto es legítimo solo aproximadamente: desaprovecha la información que contiene el orden.
- El tratamiento riguroso pertenece a *econometría de series temporales* (otra asignatura).

Series temporales

Las *series temporales* son listas de números indexadas por el tiempo. El índice t identifica un momento concreto: un año, un trimestre, un mes, un día, incluso un segundo. Y aquí, a diferencia de la sección cruzada, *el orden del índice sí contiene información esencial*: el dato de febrero de 2020 viene después del de enero de 2020, y esa relación temporal es parte de la naturaleza del fenómeno. No tiene sentido reordenar los meses: si lo hiciéramos, perderíamos la información de que la pandemia de COVID-19 estalló en marzo de 2020 y sus efectos se sintieron inmediatamente después, de que el verano sigue a la primavera, o de que una recesión empieza antes de tocar fondo.

Formalmente, una serie temporal sigue siendo un vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, pero ese vector lleva ahora información *adicional* codificada en el orden de sus componentes. La estructura matemática de $\mathbb{R}^n$ que usaremos en el curso (productos escalares, normas, proyecciones) *no aprovecha* esa información temporal.

Esto plantea una decisión metodológica importante. En este curso vamos a tratar las series temporales *como si fueran* datos de sección cruzada. Esto significa que, cuando trabajemos con la tasa de paro mensual española de los últimos veinte años, calcularemos medias, varianzas y regresiones con las mismas fórmulas que aplicaríamos a una muestra de trabajadores observados en un mismo instante. Esta simplificación es *legítima solo aproximadamente*: desaprovecha la información que contiene el orden de las observaciones (las cercanas en el tiempo tienden a parecerse más entre sí que las lejanas), y eso tiene consecuencias sobre la inferencia estadística —los errores estándar, los contrastes de hipótesis— que este curso no está en condiciones de tratar. El tratamiento riguroso pertenece a una asignatura específica de *econometría de series temporales*.

Aun así, las series temporales son omnipresentes en macroeconomía y en economía financiera, y aprender a manejarlas — aunque sea con las limitaciones de este curso — es indispensable. Muchos de los datasets que vienen con Gretl son series temporales: PIB trimestral, tipos de interés, índices bursátiles, agregados monetarios. Aprenderemos a importarlos, describirlos y modelarlos con las herramientas del curso, conscientes de que estamos dejando información sobre la mesa.

Para profundizar:

- Stock, J. H. y Watson, M. W. (2019). *Introduction to Econometrics* (4ª ed.), capítulos 15 y 16. Introducción accesible a series temporales en un manual general.
- Banco de España: estadísticas y series temporales. Indicadores macroeconómicos y financieros españoles.
- INE: tasa de paro mensual. Serie con la que ejemplificaremos en clase.
- FRED (Federal Reserve Bank of St. Louis): base de datos macroeconómica internacional. Más de 800.000 series, descargables.

7 Datos de panel

- Cada observación tiene *dos índices*: (i, t).
  - i: individuo, país, empresa, etc.
  - t: instante de tiempo.
- Ejemplo: PIB de cada país de la UE en cada año desde 2000.
- Muy potentes: permiten controlar características no observadas pero que son constantes en el tiempo.
- Requieren técnicas específicas (efectos fijos, efectos aleatorios).

No los trataremos en este curso.

Datos de panel

Los *datos de panel* (o *longitudinales*) combinan las dos dimensiones anteriores: observamos las mismas unidades a lo largo del tiempo. Cada observación tiene por tanto *dos* índices: uno para la unidad (i) y otro para el tiempo (t). Si seguimos a N países durante T años, tendremos en total NT observaciones organizadas en una estructura bidimensional.

El ejemplo clásico es el seguimiento del PIB, la inflación y otras variables macroeconómicas para los países de la UE durante varias décadas. Pero también son datos de panel los estudios que siguen a los mismos individuos a lo largo de varios años (por ejemplo, la *Encuesta de Condiciones de Vida* en su versión longitudinal, o el *Panel Study of Income Dynamics* estadounidense), o el seguimiento de empresas a lo largo del tiempo en bases de datos como *Compustat* o *SABI*.

La gran ventaja de los datos de panel es que permiten *controlar características no observadas de cada unidad* siempre que esas características permanezcan constantes en el tiempo. Por ejemplo, si queremos estimar el efecto de una política fiscal sobre el crecimiento, los datos de panel nos permiten comparar a cada país *consigo mismo* antes y después de la política, eliminando así el efecto de todos los factores idiosincrásicos de ese país (geografía, cultura, historia) que de otro modo contaminarían la comparación. Esta es una herramienta extraordinariamente potente para la inferencia causal, y es una de las razones por las que los datos de panel han ganado tanto protagonismo en la economía empírica moderna.

A cambio, los datos de panel requieren técnicas específicas (efectos fijos, efectos aleatorios, métodos dinámicos como GMM), que extienden las ideas básicas de la regresión pero introducen complicaciones que no podemos abordar en un curso introductorio. *No los trataremos en esta asignatura.* Se los menciono para que sepa que existen, los reconozca cuando aparezcan en lecturas y entienda por qué muchos artículos empíricos modernos se basan en este tipo de datos. Quien siga con economía o políticas públicas los verá en cursos posteriores.

Para profundizar:

- Wooldridge, J. M. (2020), capítulos 13 y 14. Introducción a panel desde un curso introductorio.
- Vídeo: "What is Panel Data?", de Ben Lambert. Tres minutos.

- Eurostat: bases de datos longitudinales europeas. Ejemplos reales accesibles.

8 Tabla resumen de la tipología de datos

Tipo	Índice	¿Orden informativo?	En este curso
Sección cruzada	i	No	Sí (foco principal)
Serie temporal	t	Sí	Sí, como s. cruzada
Panel	(i, t)	Solo en t	No

Mensaje: *el mismo objeto matemático (un vector de $\mathbb{R}^n$) puede llevar muy distinta información según qué signifique su índice.*

Tabla resumen de la tipología de datos

La tabla de esta transparencia condensa las distinciones anteriores. Conviene leerla con un mensaje en mente: el *objeto matemático* — un vector de $\mathbb{R}^n$ — es el mismo en todos los casos, pero su *interpretación* depende crucialmente de qué significa el índice.

Si le proporcionan un vector de datos, no podrá decirme de qué tipo son hasta que sepa qué es el índice. Si el índice es una persona, podría ser el salario por hora de 240 trabajadores: un dato de sección cruzada, susceptible de un análisis directo con nuestras herramientas. Si el índice es un mes, podría ser la tasa de paro mensual de España durante 20 años: una serie temporal, que trataremos con las cautelas ya señaladas. Si el índice combina país y año, sería un panel, que queda fuera del alcance del curso.

Esta diferencia *no se ve* en los números: se ve en el contexto. Y es responsabilidad del analista — no del software, no del algoritmo — saber qué tiene entre manos antes de aplicar técnica alguna. Este es uno de los hilos que recorrerán el curso: las técnicas matemáticas son ciegas a la interpretación, y por eso quien las aplica debe ver lo que ellas no ven.

Una nota final sobre nomenclatura. En la literatura inglesa encontrará las expresiones *cross-sectional data*, *time series* y *panel data* (o *longitudinal data*). En castellano se usan *datos de sección cruzada* (a veces *datos de corte transversal*), *series temporales* (o *series de tiempo*) y *datos de panel* (o *datos longitudinales*). Las usaremos indistintamente.

Para profundizar:

- Wooldridge, J. M. (2020), capítulo 1, sección 1.3. Tabla resumen y ejemplos en el manual de referencia.
- Cameron, A. C. y Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*, capítulo 1. Para una visión más avanzada, cuando llegue el momento.

9 Cómo trabajaremos durante el curso

- *Teoría* con intuición geométrica explícita:
  - regresión = proyección ortogonal en $\mathbb{R}^n$.
- *Laboratorio* con Gretl:

  - datos *reales* → enfrentarse a la complejidad empírica.
  - datos *simulados* → verificar que la teoría funciona.
- Cada sesión: *test corto* (15–20 min) al final.
- Regla pedagógica:

Preferimos entender bien pocas cosas que entender mal muchas.

Cómo trabajaremos durante el curso

Conviene cerrar la sesión con una explicación del método de trabajo. Las decisiones que voy a explicar no son arbitrarias: responden a un diagnóstico sobre qué funciona y qué no en un curso introductorio de econometría con un alumnado tan heterogéneo como este.

Las *sesiones de teoría* introducen los conceptos con una intuición geométrica explícita. La regresión lineal, que es el corazón del curso, se entenderá como una *proyección ortogonal* en $\mathbb{R}^n$: dado un vector de datos (la variable que queremos explicar) y un conjunto de vectores explicativos, buscamos la combinación lineal de estos últimos que mejor aproxima al primero. "Mejor" se definirá geométricamente, como la combinación cuyo error tiene la norma más pequeña. Verá que esta visión, aunque al principio resulte abstracta, da una coherencia notable a todo el curso: la media, la varianza, la regresión simple, la regresión múltiple, el R^2 y la correlación se entienden como caras de una misma idea geométrica.

Las *sesiones de laboratorio* con Gretl son la otra cara del curso. Las dedicaremos a trabajar con dos tipos de datos:

Por un lado, *datos reales* — datasets que vienen con Gretl o que descargaremos de fuentes públicas. Trabajar con datos reales es indispensable porque nos enfrenta a la complejidad de los problemas empíricos: variables mal medidas, datos faltantes, decisiones discutibles sobre qué incluir y qué excluir, resultados que no encajan con la teoría. El analista en formación tiene que acostumbrarse a esta complejidad: en el mundo real no hay respuestas únicas y limpias.

Por otro lado, *datos simulados* — datos que generaremos nosotros mismos con propiedades conocidas. Las simulaciones son una herramienta pedagógica extraordinariamente potente, infrautilizada en muchos cursos. Permiten *verificar* que la teoría funciona como prometemos: si decimos que el estimador MCO es insesgado bajo ciertos supuestos, podemos simular miles de muestras que cumplan esos supuestos, calcular el estimador en cada una y comprobar que su promedio coincide con el verdadero parámetro. Esta verificación empírica de la teoría es, en mi opinión, una de las experiencias más formativas que un curso de econometría puede ofrecer.

En casi todas las sesiones reservaremos los últimos *15 a 20 minutos para un ejercicio tipo test*. No es un instrumento de evaluación punitiva: es un diagnóstico. Sirve para que usted detecte a tiempo qué conceptos no han quedado claros, y para que yo detecte qué partes de la explicación necesitan reforzarse. Le animo a tomarlo como una oportunidad de aprendizaje activo, no como un examen.

Cierro con la regla pedagógica del curso, que es honesta y conviene tener presente: *preferimos comprender bien unas pocas cosas que malentender muchas*. Si en algún momento siente que vamos demasiado deprisa o que un concepto no ha quedado claro, *dígalo*. Es preferible recortar contenido

— y lo recortaremos, sin dramatizar — que avanzar sobre una base que no se ha asentado. La econometría no es un catálogo de técnicas que se memorizan: es una forma de pensar sobre datos económicos, y esa forma de pensar se construye despacio.

Para profundizar:

- Kennedy, P. (2008). *A Guide to Econometrics* (6ª ed.). Manual conversacional, sin formalismo excesivo, ideal para acompañar el curso.
- Sobre el uso pedagógico de simulaciones: Kennedy, P. (1998). *Using Monte Carlo studies for teaching econometrics*, en *Teaching Undergraduate Econometrics*. Argumenta lo que aquí he resumido.
- Adkins, L. C. (2018). *Using Gretl for Principles of Econometrics*. Manual gratuito que acompaña al texto de Hill, Griffiths y Lim. Útil como referencia práctica.

10 Próxima sesión

- Vectores en $\mathbb{R}^2$ y $\mathbb{R}^3$: puntos en el espacio, suma, producto por escalar.
- Salto a $\mathbb{R}^n$: listas de números que se manejan igual (operaciones componente a componente).
- Producto escalar y norma.
- *Antes de la sesión*: descargar e instalar Gretl (`http://gretl.sourceforge.net`).

Próxima sesión

La próxima sesión es enteramente matemática: no veremos econometría todavía, sino el lenguaje geométrico en el que la formularemos durante el resto del curso.

Empezaremos con vectores en $\mathbb{R}^2$ y $\mathbb{R}^3$, donde podemos dibujarlos como flechas y manipularlos con la intuición visual del bachillerato: sumar dos vectores poniéndolos cabeza con cola, multiplicar un vector por un escalar alargándolo o acortándolo. Una vez asentada esa intuición, daremos el salto a $\mathbb{R}^n$: no como un objeto exótico, sino simplemente como *listas de n números que se manejan igual* que las listas de 2 o 3 números, solo que sin posibilidad de dibujo.

Introduciremos entonces dos operaciones que serán las protagonistas del curso: el *producto escalar* y la *norma*. El producto escalar de dos vectores es la suma de los productos de sus componentes correspondientes; mide, en cierto sentido, hasta qué punto dos vectores "apuntan en la misma dirección". La norma de un vector es la raíz cuadrada del producto escalar consigo mismo; mide su "longitud". Veremos que estas dos nociones bastan para definir geométricamente la *ortogonalidad* (cuando el producto escalar es cero) y el *ángulo* entre dos vectores, y que con ellas se construye todo el edificio del curso.

Un encargo concreto antes de la próxima sesión de laboratorio: *instale Gretl* en su ordenador. Es gratuito, ocupa muy poco espacio y funciona en Windows, macOS y Linux. Lo puede descargar en http://gretl.sourceforge.net para usarlo en su propio ordenador. Los puestos de las salas de ordenadores de la facultad también tienen Gretl instalado.

Si quiere ir abriendo boca, en la próxima sesión recomendaré algunos vídeos y lecturas concretas sobre vectores; pero si ya quiere adelantar algo, los dos primeros capítulos de la serie Essence of Linear Algebra de 3Blue1Brown son una inversión de 25 minutos extraordinariamente bien empleada. También puede ver los cuatro vídeos que he elaborado para mis alumnos del grado de economía: Curso de Álgebra Lineal,