



# IN4402 - Aplicaciones de Probabilidades y Estadística en Gestión

Ronald Lelebici - Angelo Muñoz

`ronald.lelebici@gmail.com`

`gelox97@gmail.com`

Facultad de Ciencias Físicas y Matemáticas  
Universidad de Chile

25 de septiembre del 2018



# ¿Qué estudiaremos hoy?

## P1. Comente

Teorías y evidencias

Causalidad y correlación

Tipos de bases de datos

## P2. Formas funcionales

## P3. Modelamiento e interpretación

## P4. Estimación computacional de MCO



## P1.1. Teorías y evidencias

**[P1, Control 1, 2017]** Explique si la acumulación de evidencia empírica que no contradice las predicciones de una teoría fortalece a esta última.



## P1.1. Teorías y evidencias

**[P1, Control 1, 2017]** Explique si la acumulación de evidencia empírica que no contradice las predicciones de una teoría fortalece a esta última.

### **Respuesta:**

La acumulación de evidencia empírica que no contradice a una teoría la fortalecerá siempre que:

- No sea explicada de mejor manera por otra teoría (siguiendo el pensamiento de Thomas Kuhn).
- Dicha evidencia proporcione nueva información que permita generalizar resultados, en lugar de ser acotada y redundante.



## P1.2. Causalidad y correlación: salud mental

**[P2, Control 1, 2017]** En un sitio web se afirma que: *Un estudio conjunto de la Universidad de Warwick y Manchester concluye que estar rodeados de amigos mentalmente saludables ayuda a nuestra propia salud mental.*



## P1.2. Causalidad y correlación: salud mental

**[P2, Control 1, 2017]** En un sitio web se afirma que: *Un estudio conjunto de la Universidad de Warwick y Manchester concluye que estar rodeados de amigos mentalmente saludables ayuda a nuestra propia salud mental. Los investigadores explican que las personas que están habitualmente de buen humor a nuestro lado, nos contagian de ese estado de ánimo.*



## P1.2. Causalidad y correlación: salud mental

**[P2, Control 1, 2017]** En un sitio web se afirma que: *Un estudio conjunto de la Universidad de Warwick y Manchester concluye que estar rodeados de amigos mentalmente saludables ayuda a nuestra propia salud mental. Los investigadores explican que las personas que están habitualmente de buen humor a nuestro lado, nos contagian de ese estado de ánimo.*

Explique qué dificultades podrían haber para probar que salud mental de una persona **es causada** por la de sus amigos.



## P1.2. Causalidad y correlación: salud mental

**[P2, Control 1, 2017]** En un sitio web se afirma que: *Un estudio conjunto de la Universidad de Warwick y Manchester concluye que estar rodeados de amigos mentalmente saludables ayuda a nuestra propia salud mental. Los investigadores explican que las personas que están habitualmente de buen humor a nuestro lado, nos contagian de ese estado de ánimo.*

Explique qué dificultades podrían haber para probar que salud mental de una persona **es causada** por la de sus amigos.

### Respuesta:

A continuación se enuncian posibles dificultades:



## P1.2. Causalidad y correlación: salud mental

[P2, Control 1, 2017] En un sitio web se afirma que: *Un estudio conjunto de la Universidad de Warwick y Manchester concluye que estar rodeados de amigos mentalmente saludables ayuda a nuestra propia salud mental. Los investigadores explican que las personas que están habitualmente de buen humor a nuestro lado, nos contagian de ese estado de ánimo.*

Explique qué dificultades podrían haber para probar que salud mental de una persona **es causada** por la de sus amigos.

### Respuesta:

A continuación se enuncian posibles dificultades:

- 1 Defender que se trata de una causalidad y no únicamente correlación. Se deben encontrar argumentos teóricos que respalden la relación causal entre tener amigos saludables mentalmente y estar sanos en ese sentido, al igual que ellos.



## P1.2. Causalidad y correlación: salud mental

[P2, Control 1, 2017] En un sitio web se afirma que: *Un estudio conjunto de la Universidad de Warwick y Manchester concluye que estar rodeados de amigos mentalmente saludables ayuda a nuestra propia salud mental. Los investigadores explican que las personas que están habitualmente de buen humor a nuestro lado, nos contagian de ese estado de ánimo.*

Explique qué dificultades podrían haber para probar que salud mental de una persona **es causada** por la de sus amigos.

### Respuesta:

A continuación se enuncian posibles dificultades:

- 1 Defender que se trata de una causalidad y no únicamente correlación. Se deben encontrar argumentos teóricos que respalden la relación causal entre tener amigos saludables mentalmente y estar sanos en ese sentido, al igual que ellos.
- 2 Encontrar o desarrollar alguna métrica cuantitativa con sentido que mida la salud mental de las personas.



# Causalidad y correlación: memes (1)

MISLEADING STATISTICS

**OVER 2 MILLION AMERICANS  
EXPOSED TO DRINKING WATER WILL  
DIE THIS YEAR**



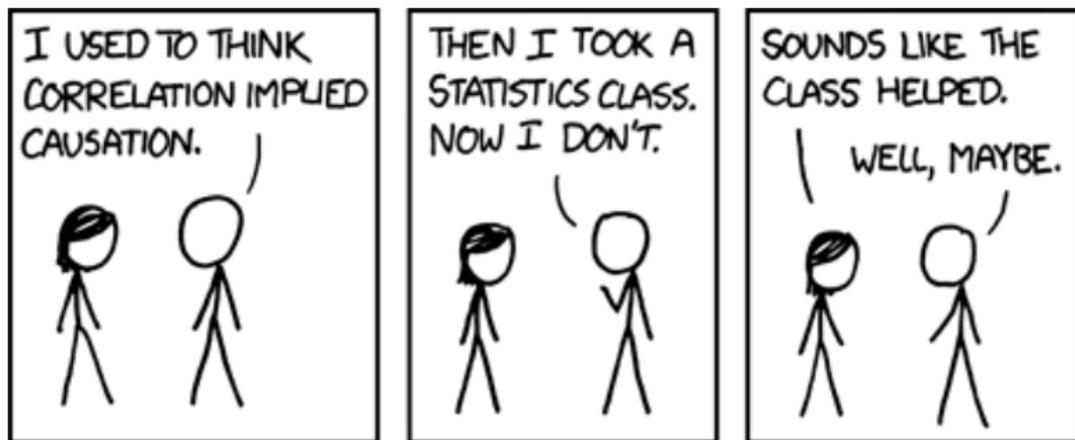
Correlation  $\neq$  Causation

CONTENT SHOULD BE USEFUL, NOT JUST PRETTY  
*vert.ms/Baddata*





## Causalidad y correlación: memes (2)





## P1.3. Causalidad y correlación: embarazo y tabaquismo

**[P1.c, Control 1, 2016]** Un médico le indica a usted que “la correlación negativa entre número de cigarrillos fumados al día durante el embarazo de una mujer y el peso de su bebé al nacer es evidencia clara del efecto negativo del tabaquismo en la salud del recién nacido”. Comente.



## P1.3. Causalidad y correlación: embarazo y tabaquismo

**[P1.c, Control 1, 2016]** Un médico le indica a usted que “la correlación negativa entre número de cigarrillos fumados al día durante el embarazo de una mujer y el peso de su bebé al nacer es evidencia clara del efecto negativo del tabaquismo en la salud del recién nacido”. Comente.

**Respuesta:** Aunque, correlación no es causalidad, hay tres posibilidades consistentes con una correlación negativa descrita:

- Más cigarrillos causan reducción del peso del bebé. ( $X \rightarrow Y$ )



## P1.3. Causalidad y correlación: embarazo y tabaquismo

**[P1.c, Control 1, 2016]** Un médico le indica a usted que “la correlación negativa entre número de cigarrillos fumados al día durante el embarazo de una mujer y el peso de su bebé al nacer es evidencia clara del efecto negativo del tabaquismo en la salud del recién nacido”. Comente.

**Respuesta:** Aunque, correlación no es causalidad, hay tres posibilidades consistentes con una correlación negativa descrita:

- Más cigarrillos causan reducción del peso del bebé. ( $X \rightarrow Y$ )
- Los bebés de menor peso causan un mayor número de cigarrillos fumados ( $Y \rightarrow X$ )



## P1.3. Causalidad y correlación: embarazo y tabaquismo

**[P1.c, Control 1, 2016]** Un médico le indica a usted que “la correlación negativa entre número de cigarrillos fumados al día durante el embarazo de una mujer y el peso de su bebé al nacer es evidencia clara del efecto negativo del tabaquismo en la salud del recién nacido”. Comente.

**Respuesta:** Aunque, correlación no es causalidad, hay tres posibilidades consistentes con una correlación negativa descrita:

- Más cigarrillos causan reducción del peso del bebé. ( $X \rightarrow Y$ )
- Los bebés de menor peso causan un mayor número de cigarrillos fumados ( $Y \rightarrow X$ )
- Existen otras variables  $Z$  tales como bajo nivel de educación de salud sobre efectos del tabaquismo en el bebé y su madre, así como pobreza. Estos factores pueden causar, a su vez, a mala nutrición de la madre, insuficientes controles médicos, etc que pueden causar, al mismo tiempo, mayor tabaquismo y menor peso del bebé ( $Y \leftarrow Z \rightarrow X$ ).



## P1.4. Corte transversal, series de tiempo y datos de panel

**[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.



## P1.4. Corte transversal, series de tiempo y datos de panel

**[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.

### Respuesta:

- Datos de corte transversal: datos de distinta índole en un momento del tiempo. **Varios individuos, un instante.**



## P1.4. Corte transversal, series de tiempo y datos de panel

**[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.

### Respuesta:

- Datos de corte transversal: datos de distinta índole en un momento del tiempo. **Varios individuos, un instante.**  
Ejemplo: PIB de países latinoamericanos en 2013.



## P1.4. Corte transversal, series de tiempo y datos de panel

**[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.

### Respuesta:

- Datos de corte transversal: datos de distinta índole en un momento del tiempo. **Varios individuos, un instante.**  
Ejemplo: PIB de países latinoamericanos en 2013.
- Series de tiempo: datos para una unidad (individuo, empresa, país, etc) a través del tiempo. **Un individuo, varios instantes.**



## P1.4. Corte transversal, series de tiempo y datos de panel

**[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.

### Respuesta:

- Datos de corte transversal: datos de distinta índole en un momento del tiempo. **Varios individuos, un instante.** Ejemplo: PIB de países latinoamericanos en 2013.
- Series de tiempo: datos para una unidad (individuo, empresa, país, etc) a través del tiempo. **Un individuo, varios instantes.** Ejemplo: PIB anual para Chile 1990-2013.



## P1.4. Corte transversal, series de tiempo y datos de panel

**[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.

### Respuesta:

- Datos de corte transversal: datos de distinta índole en un momento del tiempo. **Varios individuos, un instante.** Ejemplo: PIB de países latinoamericanos en 2013.
- Series de tiempo: datos para una unidad (individuo, empresa, país, etc) a través del tiempo. **Un individuo, varios instantes.** Ejemplo: PIB anual para Chile 1990-2013.
- Datos de panel: datos de distinta índole a través de diferentes momentos del tiempo. **Varios individuos, varios instantes.**



## P1.4. Corte transversal, series de tiempo y datos de panel

**[P1.b, Control 1, 2016]** Explique qué son datos de corte transversal, series de tiempo y datos de panel. Dé un ejemplo para cada uno de ellos.

### Respuesta:

- Datos de corte transversal: datos de distinta índole en un momento del tiempo. **Varios individuos, un instante.** Ejemplo: PIB de países latinoamericanos en 2013.
- Series de tiempo: datos para una unidad (individuo, empresa, país, etc) a través del tiempo. **Un individuo, varios instantes.** Ejemplo: PIB anual para Chile 1990-2013.
- Datos de panel: datos de distinta índole a través de diferentes momentos del tiempo. **Varios individuos, varios instantes.** Ejemplo: PIB anual de países latinoamericanos 1980-2013.



# Datos de corte transversal

Contienen información de distintos **individuos**.

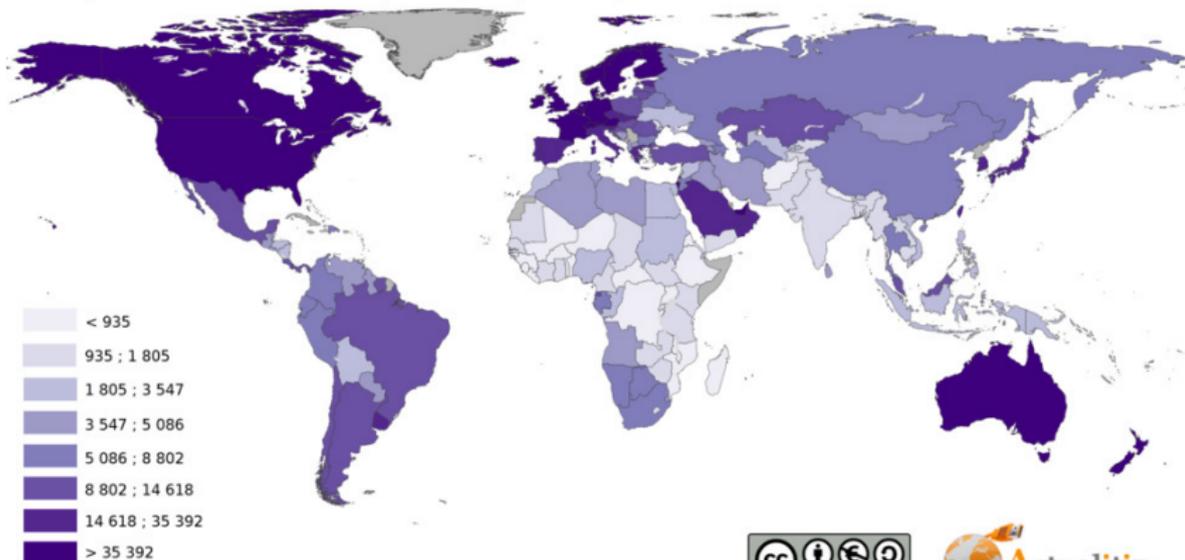
	edad	escolaridad	ln_ingreso	genero
1	15	10	.	1
2	45	8	13.33133	1
3	55	8	12.69645	1
4	10	.	.	1
5	19	15	.	0
6	10	.	.	0
7	23	12	.	1
8	30	5	12.11355	1
9	22	14	.	0
10	10	.	.	0
11	33	12	12.52906	1
12	63	6	.	1
13	3	.	.	1
14	67	3	10.91962	0



# Datos de corte transversal

Contienen información de distintos **individuos**.

El producto interior bruto per capita (Dólares)



Fuente : FMI - 2015

Copyright © Actualitix.com All rights reserved





## Series de tiempo

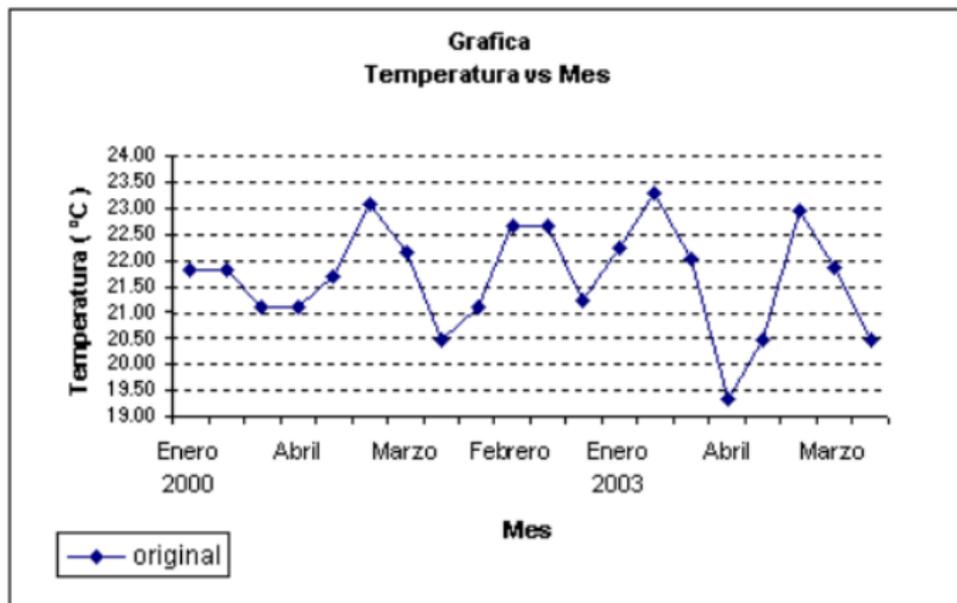
Contienen información de **un individuo** en distintos **periodos**.

Semana	Ventas
1	17
2	21
3	19
4	23
5	18
6	16
7	20
8	18
9	22
10	20
11	15
12	22



# Series de tiempo

Contienen información de **un individuo** en distintos **periodos**.





## Datos de panel

Contienen información de distintos **individuos** en distintos **periodos**.

<i>individuo</i>	<i>periodo</i>	<i>ingresos</i>	<i>edad</i>	<i>sexo</i>
1	2003	1500	27	1
1	2004	1700	28	1
1	2005	2000	29	1
2	2003	2100	41	2
2	2004	2100	42	2
2	2005	2200	43	2

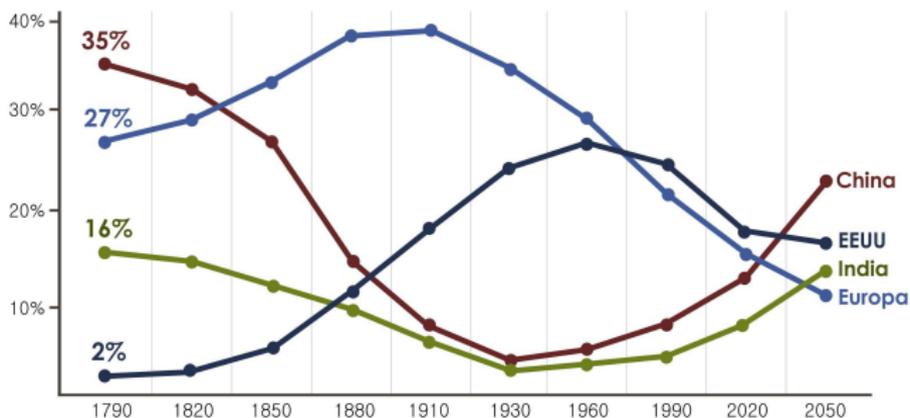


# Datos de panel

Contienen información de distintos **individuos** en distintos **periodos**.

Porcentaje del PIB global

Banco de Datos  
elOrdenMundial.com



fuelle: Angus Maddison, Fondo Monetario Internacional



## P2. Lin-lin, log-log, lin-log y log-lin

Suponiendo causalidad, demuestre que:

- 1 En un modelo lin-lin ( $Y_i = \beta_0 + \beta_1 X_i + U_i$ ), si  $X$  aumenta en una unidad,  $Y$  aumenta en  $\beta_1$  unidades.



## P2. Lin-lin, log-log, lin-log y log-lin

Suponiendo causalidad, demuestre que:

- 1 En un modelo lin-lin ( $Y_i = \beta_0 + \beta_1 X_i + U_i$ ), si  $X$  aumenta en una unidad,  $Y$  aumenta en  $\beta_1$  unidades.
- 2 En un modelo log-log ( $\ln Y_i = \beta_0 + \beta_1 \ln X_i + U_i$ ), si  $X$  cambia en un 1%,  $Y$  cambia en  $\beta_1$  %.



## P2. Lin-lin, log-log, lin-log y log-lin

Suponiendo causalidad, demuestre que:

- 1 En un modelo lin-lin ( $Y_i = \beta_0 + \beta_1 X_i + U_i$ ), si  $X$  aumenta en una unidad,  $Y$  aumenta en  $\beta_1$  unidades.
- 2 En un modelo log-log ( $\ln Y_i = \beta_0 + \beta_1 \ln X_i + U_i$ ), si  $X$  cambia en un 1%,  $Y$  cambia en  $\beta_1$  %.
- 3 En un modelo lin-log ( $Y_i = \beta_0 + \beta_1 \ln X_i + U_i$ ), si  $X$  cambia en un 1%,  $Y$  cambia en  $\frac{\beta_1}{100}$  unidades.



## P2. Lin-lin, log-log, lin-log y log-lin

Suponiendo causalidad, demuestre que:

- 1 En un modelo lin-lin ( $Y_i = \beta_0 + \beta_1 X_i + U_i$ ), si  $X$  aumenta en una unidad,  $Y$  aumenta en  $\beta_1$  unidades.
- 2 En un modelo log-log ( $\ln Y_i = \beta_0 + \beta_1 \ln X_i + U_i$ ), si  $X$  cambia en un 1%,  $Y$  cambia en  $\beta_1$  %.
- 3 En un modelo lin-log ( $Y_i = \beta_0 + \beta_1 \ln X_i + U_i$ ), si  $X$  cambia en un 1%,  $Y$  cambia en  $\frac{\beta_1}{100}$  unidades.
- 4 En un modelo log-lin ( $\ln Y_i = \beta_0 + \beta_1 X_i + U_i$ ), si  $X$  cambia en una unidad,  $Y$  cambia en  $100 \cdot \beta_1$  %.



## P3. Precios de pasajes aéreos

Usted tiene un flamante trabajo en la industria de transporte aéreo. Su primera tarea será construir modelos econométricos que expliquen los precios ofrecidos en la industria. Considere que tenemos un modelo para explicar precios efectivos que se cobran por pasajes aéreos de ida y regreso. Formule un modelo que permita estimar:

- 1 Efecto porcentual esperado en el precio del día de la semana en que el pasaje se emite y la dif. porcentual aprox. 2 días de la semana.



## P3. Precios de pasajes aéreos

Usted tiene un flamante trabajo en la industria de transporte aéreo. Su primera tarea será construir modelos econométricos que expliquen los precios ofrecidos en la industria. Considere que tenemos un modelo para explicar precios efectivos que se cobran por pasajes aéreos de ida y regreso. Formule un modelo que permita estimar:

- 1 Efecto porcentual esperado en el precio del día de la semana en que el pasaje se emite y la dif. porcentual aprox. 2 días de la semana.
- 2 Efecto porcentual del tiempo de estadía (tiempo entre el viaje de ida y de vuelta), considerando la posible importancia del día de la semana en que se viaja. Por ejemplo: no es lo mismo un viaje viernes-domingo, que uno martes-jueves.



## P3. Precios de pasajes aéreos

Usted tiene un flamante trabajo en la industria de transporte aéreo. Su primera tarea será construir modelos econométricos que expliquen los precios ofrecidos en la industria. Considere que tenemos un modelo para explicar precios efectivos que se cobran por pasajes aéreos de ida y regreso. Formule un modelo que permita estimar:

- 1 Efecto porcentual esperado en el precio del día de la semana en que el pasaje se emite y la dif. porcentual aprox. 2 días de la semana.
- 2 Efecto porcentual del tiempo de estadía (tiempo entre el viaje de ida y de vuelta), considerando la posible importancia del día de la semana en que se viaja. Por ejemplo: no es lo mismo un viaje viernes-domingo, que uno martes-jueves.
- 3 La respuesta porcentual del precio a la distancia entre origen y destino, considerando diferencias en dicha respuesta al n de escalas.



## P3. Precios de pasajes aéreos

Usted tiene un flamante trabajo en la industria de transporte aéreo. Su primera tarea será construir modelos econométricos que expliquen los precios ofrecidos en la industria. Considere que tenemos un modelo para explicar precios efectivos que se cobran por pasajes aéreos de ida y regreso. Formule un modelo que permita estimar:

- 1 Efecto porcentual esperado en el precio del día de la semana en que el pasaje se emite y la dif. porcentual aprox. 2 días de la semana.
- 2 Efecto porcentual del tiempo de estadía (tiempo entre el viaje de ida y de vuelta), considerando la posible importancia del día de la semana en que se viaje. Por ejemplo: no es lo mismo un viaje viernes-domingo, que uno martes-jueves.
- 3 La respuesta porcentual del precio a la distancia entre origen y destino, considerando diferencias en dicha respuesta al n de escalas.
- 4 Efecto porcentual en el precio del tiempo con anticipación de venta del pasaje, permitiendo **la posibilidad (no imponiendo)** un reacción nula para viajes comprados con 6 meses o más de anticipación.



## P3. Precios de pasajes aéreos

Usted tiene un flamante trabajo en la industria de transporte aéreo. Su primera tarea será construir modelos econométricos que expliquen los precios ofrecidos en la industria. Considere que tenemos un modelo para explicar precios efectivos que se cobran por pasajes aéreos de ida y regreso. Formule un modelo que permita estimar:

- 1 Efecto porcentual esperado en el precio del día de la semana en que el pasaje se emite y la dif. porcentual aprox. 2 días de la semana.
- 2 Efecto porcentual del tiempo de estadía (tiempo entre el viaje de ida y de vuelta), considerando la posible importancia del día de la semana en que se viaje. Por ejemplo: no es lo mismo un viaje viernes-domingo, que uno martes-jueves.
- 3 La respuesta porcentual del precio a la distancia entre origen y destino, considerando diferencias en dicha respuesta al n de escalas.
- 4 Efecto porcentual en el precio del tiempo con anticipación de venta del pasaje, permitiendo **la posibilidad (no imponiendo)** un reacción nula para viajes comprados con 6 meses o más de anticipación.



## P4. Estimación computacional de MCO

Haremos la estimación de MCO de tres formas distintas:

- 1 **Stata:** forma “fácil”. El software hace la pega por nosotros.
- 2 **Matlab:** usaremos el resultado de la minimización de la suma de los residuos al cuadrado, cuya solución es el estimador  $\hat{\beta} = (X^T X)^{-1} X^T Y$ . Nosotros mismos haremos las operaciones matriciales.
- 3 **Excel:** mediante el *solver*, optimizaremos la suma de los residuos al cuadrado usando métodos numéricos (no usaremos la fórmula de  $\hat{\beta}$ ).



## ¿Qué aprendimos?

- 1 **Teorías y evidencias:** analizamos parte del pensamiento que expone Thomas Kuhn en “La estructura de las revoluciones científicas”. La acumulación de evidencia que no contradice una teoría la fortalecerá siempre que: (i) no sea explicada mejor por otra teoría y (ii) permita generalizar resultados.



## ¿Qué aprendimos?

- 1 **Teorías y evidencias:** analizamos parte del pensamiento que expone Thomas Kuhn en “La estructura de las revoluciones científicas”. La acumulación de evidencia que no contradice una teoría la fortalecerá siempre que: (i) no sea explicada mejor por otra teoría y (ii) permita generalizar resultados.
- 2 **Correlación y causalidad:** debemos ser cuidadosos y no confundirlas. Debemos tener una teoría fundada para atribuir efectos de causalidad a correlaciones numéricas.



## ¿Qué aprendimos?

- 1 **Teorías y evidencias:** analizamos parte del pensamiento que expone Thomas Kuhn en “La estructura de las revoluciones científicas”. La acumulación de evidencia que no contradice una teoría la fortalecerá siempre que: (i) no sea explicada mejor por otra teoría y (ii) permita generalizar resultados.
- 2 **Correlación y causalidad:** debemos ser cuidadosos y no confundirlas. Debemos tener una teoría fundada para atribuir efectos de causalidad a correlaciones numéricas.
- 3 **Formas funcionales:** distintos tipos de modelos pueden explicar mejor o peor el comportamiento de ciertos fenómenos. La interpretación de los coeficientes  $\beta_k$  que estimemos dependerá de cómo construyamos este modelo. Derivando podemos inferir su significado.



## ¿Qué aprendimos?

- 1 Teorías y evidencias:** analizamos parte del pensamiento que expone Thomas Kuhn en “La estructura de las revoluciones científicas”. La acumulación de evidencia que no contradice una teoría la fortalecerá siempre que: (i) no sea explicada mejor por otra teoría y (ii) permita generalizar resultados.
- 2 Correlación y causalidad:** debemos ser cuidadosos y no confundirlas. Debemos tener una teoría fundada para atribuir efectos de causalidad a correlaciones numéricas.
- 3 Formas funcionales:** distintos tipos de modelos pueden explicar mejor o peor el comportamiento de ciertos fenómenos. La interpretación de los coeficientes  $\beta_k$  que estimemos dependerá de cómo construyamos este modelo. Derivando podemos inferir su significado.
- 4 Estimación de MCO:** es algo tangible. No está solo en el papel. Stata no es una caja negra. Ahora sabemos concretamente cómo se realiza.