



# Proyecto de curso

## Descripción

El proyecto final tiene como objetivo que los estudiantes apliquen los conocimientos adquiridos en el curso para abordar un problema real utilizando técnicas de análisis de datos y análisis de regresión.

El proyecto se llevará a cabo en grupos de 3 a 6 personas.

De este se tienen dos entregables, una presentación final y un notebook que documente el análisis realizado.

## Instrucciones Generales:

Cada grupo debe seleccionar un conjunto de datos y proponer una pregunta de investigación y una hipótesis a testear, que puedan desarrollar mediante análisis de regresión.

## Estructura del proyecto

---

Se les recomienda considerar la siguiente estructura del proyecto:

1. **Introducción:** Introduce el tema, destacando su relevancia y por qué se centra en las variables específicas. Contextualiza y explica (brevemente) la importancia del problema.
2. **Preguntas e Hipótesis:** Describe la pregunta de investigación y la principal hipótesis que proponen para abordarla.
3. **Análisis Descriptivo de los Datos:** Explica los datos utilizados e incluye estadísticas descriptivas (media, varianza, mínimo, máximo) y visualizaciones como histogramas, box plot, scatter plot, correlograma, etc.
4. **Modelamiento:** Describe el modelo de regresión básico que han analizado, junto con las versiones “enriquecidas” que consideraron.
5. **Resultados de Estimación:** Presenta y explica los resultados del modelo, interpretando los coeficientes y sus implicaciones. Evalúa la hipótesis inicial propuesta con los resultados obtenidos, particularmente la significancia estadística de los coeficientes estimados.
6. **Comparación de Modelos Alternativos:** Describe cómo evaluaron y compararon diferentes especificaciones del modelo.
7. **Evaluación del Modelo:** Evalúa la validez del modelo, incluyendo pruebas de heterocedasticidad (White y/o Breusch Pagan), multicolinealidad (VIF) y discute otros posibles sesgos en la estimación, por variables omitidas.
8. **Conclusiones:** Resalta las conclusiones clave de su análisis.

9. Limitaciones del Estudio: Discute las limitaciones del enfoque y los datos utilizados.

## Entregables y evaluación

El proyecto se divide en dos entregables: una presentación final de resultados y una documentación que detalle cómo se realizaron los cálculos y análisis.

La presentación final debe abordar varios aspectos clave, incluyendo la descripción de los datos, el modelado y la evaluación del modelo, conclusiones y limitaciones.

## Fechas y modo de entrega

---

- La entrega del notebook es hasta el día 26 de octubre, por canvas. Solo debe ir el notebook en ipynb.
- Presentación final: una presentación de diapositiva de los resultados, entre 10 y 15 minutos de duración.
  - Modalidad sincrónica: se presenta en una reunion de 15 minutos con la profesora. Debe agendar en <https://doodle.com/meeting/participate/id/dympLG7d>
  - Modalidad Asincrónica: entrega un link de video subido a youtube, dropbox o plataforma afin. Esta entrega es hasta el 26 de octubre en CANVAS.

## Datos disponibles

Para el desarrollo del proyecto pueden usar cualquiera o una combinación de los siguientes sets de datos disponibles:

### 1. Datos de educacion de Chile

Cuenta con un set de datos sobre el sistema educativo chleno, que incluyen matricula universitaria, pruebas de amsisión, entre otros.

Link oficial: <https://datosabiertos.mineduc.cl> Acceder carpeta datos:  
<https://www.dropbox.com/sh/pm0rxl18gr56m80/AAB0wdX4FT3rpk4ZYxvjouSya?dl=0>

### 2. Datos socioeconómicos de Chile - Encuesta Casen

La Encuesta Casen es realizada por el Ministerio de Desarrollo Social y Familia en Chile desde la década de 1990. Su propósito principal es recopilar información que permita comprender la situación de los hogares y la población, en especial aquellos en situación de pobreza y grupos prioritarios para la política social. Esta encuesta se lleva a cabo de manera periódica y abarca una amplia variedad de temas, como ingresos, educación, empleo, salud, vivienda y otros aspectos relevantes para comprender la realidad socioeconómica de la población chilena.

[Sitio Web oficial](#)

(En ese mismi link, pueden descargar encuestas anteriores)

### 3. Datos de licitaciones y compras de comunas de Chile

En Chile, las compras publicas realizadas por las comunas se registran pvinculadas ala ley de transparencia. Contamos con datos de las órdenes de compra entre 2014 y 2022, de las cuales está presente un subset para algunas categorías de producto con la siguiente información:

```
[“ID”,“OrganismoPublico”,“sector”,“monedaItem”,“cantidad”, “precioNeto”,
“totalLineaNeto”,“CiudadUnidadCompra”,“codigoProductoONU”,
“FechaCreacion”,“FechaAceptacion”,“FechaEnvio”,“FechaCancelacion”,“ProcedenciaOC”,
“CodigoLicitacion”,“CodigoProveedor”,“NombreProveedor”]
```

Descargar datos comprimidos: [https://www.dropbox.com/scl/fi/pv7rmrtavilc595hboiw5/db\\_mel2.rar?rlkey=lvj4b2i5t7icpj0udrluth5j8&dl=0](https://www.dropbox.com/scl/fi/pv7rmrtavilc595hboiw5/db_mel2.rar?rlkey=lvj4b2i5t7icpj0udrluth5j8&dl=0)

#### 4. Datos de crecimiento y actividad económica del mundo

Mediante la API del Banco Mundial, podemos acceder a varias series temporales sobre ingreso y crecimiento de diversos países. Tienen más de 25.000 indicadores diferentes.

Link a documentación oficial: <https://datahelpdesk.worldbank.org/knowledgebase/articles/889386-developer-information-overview>

Ejemplo de uso:

Importamos los datos

```
import pandas as pd
#pandas remote data access support for calls to the World Bank Indicators API

from pandas_datareader import data, wb # para instalar: conda install pandas-datareader o p
```

Por ejemplo, importamos la lista de países disponibles

```
# Obtengamos la lista de países disponibles
countries=wb.get_countries()

#Preview primeras filas lista de países
countries[:5]
```

	iso3c	iso2c	name	region	adminregion	incomeLevel	lendingType	capitalCity	longitude	latitude
0	ABW	AW	Aruba	Latin America & Caribbean		High income	Not classified	Oranjestad	-70.0167	12.5167
1	AFE	ZH	Africa Eastern and Southern	Aggregates		Aggregates	Aggregates		NaN	NaN
2	AFG	AF	Afghanistan	South Asia	South Asia	Low income	IDA	Kabul	69.1761	34.5228
3	AFR	A9	Africa	Aggregates		Aggregates	Aggregates		NaN	NaN
4	AFW	ZI	Africa Western and Central	Aggregates		Aggregates	Aggregates		NaN	NaN

Podemos explorar los indicadores disponibles directamente desde los datos, vial al API. Por ejemplo aca revisamos los indicadores que contengan gdp. Pueden usar otros.

```
#Revisemos que indicadores hay disponibles. En este caso revisare de PIB (GDP en inglés), pero
wb.search('gdp')
```

	id	name	unit	source	sourceNote	sourceOrganization	topics
688	6.0.GDP_current	GDP (current \$)		LAC Equity Lab	GDP is the sum of gross value added by all res...	b'World Development Indicators (World Bank)'	Economy & Growth
689	6.0.GDP_growth	GDP growth (annual %)		LAC Equity Lab	Annual percentage growth rate of GDP at market...	b'World Development Indicators (World Bank)'	Economy & Growth
690	6.0.GDP_usd	GDP (constant 2005 \$)		LAC Equity Lab	GDP is the sum of gross value added by all res...	b'World Development Indicators (World Bank)'	Economy & Growth
691	6.0.GDPpc_constant	GDP per capita, PPP (constant 2011 internation...		LAC Equity Lab	GDP per capita based on purchasing power parit...	b'World Development Indicators (World Bank)'	Economy & Growth
1578	BG.GSR.NFSV.GD.ZS	Trade in services (% of GDP)		World Development Indicators	Trade in services is the sum of service export...	b'International Monetary Fund, Balance of Paym...	Economy & Growth ; Private Sector ; Trade
...	...	...	...	...	...	...	...
20286	UIS.XUNIT.GDPCAP.23.FSGOV	Initial government funding per secondary stude...		Education Statistics	Total general (local, regional and central, cu...	b'UNESCO Institute for Statistics'	
20287	UIS.XUNIT.GDPCAP.23.FSHH	Initial household funding per secondary studen...		Education Statistics	Total payments of households (pupils, students...	b'UNESCO Institute for Statistics'	
20288	UIS.XUNIT.GDPCAP.3.FSGOV	Initial government funding per upper secondary...		Education Statistics	Total general (local, regional and central, cu...	b'UNESCO Institute for Statistics'	
20289	UIS.XUNIT.GDPCAP.5T8.FSGOV	Initial government funding per tertiary studen...		Education Statistics	Total general (local, regional and central, cu...	b'UNESCO Institute for Statistics'	
20290	UIS.XUNIT.GDPCAP.5T8.FSHH	Initial household		Education Statistics	Total payments	b'UNESCO Institute for Statistics'	

id	name	unit	source	sourceNote	sourceOrganization	topics
	funding per tertiary student...			of households (pupils, students...		

585 rows × 7 columns

Con esta información, podemos descargar los datos para uno, o varios países, para un rango de datos indicando que indicador y las fechas a usar:

Por ejemplo, un data frame con los datos de Chile, entre 1980 y 2020 del indicador NY.GDP.PCAP.KD

```
# Descarguemos la data desde la API del banco mundial a un dataframe

df_GPDpc_Chile = wb.download(
    #Use the indicator attribute to identify which indicator or indicators to
    indicator='NY.GDP.PCAP.KD',
    #Use the country attribute to identify the countries you want data for
    country=['CL'],
    #Identify the first year for which you want the data, as an integer or a
    start='1980',
    #Identify the last year for which you want the data, as an integer or a s
    end=2020
)

df_GPDpc_Chile.head()
```

#### NY.GDP.PCAP.KD

country	year	
Chile	2020	12741.157507
	2019	13761.374474
	2018	13906.770558
	2017	13615.523858
	2016	13644.623261

## 5. Datos de complejidad económica

El índice de complejidad económica (ECI, por sus siglas en inglés) es una medida integral de las capacidades productivas de sistemas económicos amplios, generalmente ciudades, regiones o países. En particular, el ICE busca explicar el conocimiento acumulado en una población y que se manifiesta en las actividades económicas presentes en una ciudad, país o región. El equivalente en productos del Índice de Complejidad Económica es el Índice de Complejidad de Productos o Product Complexity Index (PCI).

Fueron desarrollados por Cesar A. Hidalgo, del MIT Media Lab, y Ricardo Hausmann, de la Escuela de Gobierno de la Universidad de Harvard. Los datos del ECI están disponibles en el Observatorio de Complejidad Económica.

Link oficial: <https://oec.world/en/rankings/eci/hs4/hs96?tab=rank> Link a paper original: <https://www.pnas.org/doi/full/10.1073/pnas.0900943106>

# Evaluación

Este proyecto se evaluará entonces en base a dos elementos, reporte de análisis y resultados (30%) y presentación final de resultados (40%)

## A. Reporte de análisis y resultados (30%)

Deben documentar su análisis de datos mediante un notebook. Este se revisará en si mismo, para fomentar las buenas prácticas y reproducibilidad de su análisis.

Rubrica de notebook reporte de análisis con los siguientes criterios:

- Entrega a tiempo
- Uso correcto del lenguaje y redacción a nivel profesional
- Orden
- Calidad de código
- Consistencia con presentación

Un detalle de los criterios los pueden ver en CANVAS.

## B. Presentación oral final de resultados (40%)

La presentación tiene un maximo de 10 minutos y se evaluará de acuerdo a los siguientes criterios:

- Entrega a tiempo
- Orden en la presentación
- Motivacion y descripcion del tema
- Pregunta de investigacion e hipotesis
- Descripcion de los datos
- Modelamiento
- Estimacion y resultados
- Conclusiones y limitaciones
- Calidad de la presentacion, material grafico y presentacion oral