



Pontificia Universidad Católica de Chile
Educación Profesional - Escuela de Ingeniería
Diplomado en Big Data y Ciencias de Datos
Minería de Datos
Relator: Sebastián Raveau

Evaluación 6 – Análisis de Regresión

Fecha de Entrega: domingo 1 de agosto

El objetivo de esta evaluación es aplicar regresión lineal para entender y predecir el Producto Interno Bruto (PIB) per cápita de un país. Para esto cuenta con información de 28 potenciales variables explicativas de distinta índole provenientes de 188 países. La descripción de las variables de la base de datos se encuentra en la siguiente página.

La evaluación consta de tres etapas:

1. Plantear hipótesis preliminares, indicando cuáles variables explicativas cree usted que podrían contribuir a explicar el PIB per cápita y cuál debiera ser su signo. Fundamente sus hipótesis en su intuición o el un análisis preliminar de los datos
2. Proponga y estime un modelo de regresión lineal que le parezca adecuado. Explique su procedimiento, indicando a grandes rasgos los pasos que siguió para llegar a su modelo final. No es necesario que reporte todos los modelos intermedios, simplemente basta con reportar el modelo final y explicar cómo llegó a él.
3. A la luz de los resultados del punto 2, discuta y analice las hipótesis planteadas en el punto 1.

Debe generar un breve reporte con su análisis y entregarlo hasta el domingo 1 de agosto al correo: mineria.datos.PUC@gmail.com. El asunto del correo y el nombre del archivo deben comenzar con [Evaluación 6] seguido con los apellidos de los estudiantes. Ejemplo: “[Evaluación 6] Gutiérrez y Soto”.

El objetivo de la tarea es demostrar que entiende lo que está haciendo R al aplicar Mínimos Cuadrados. Por favor presente sólo la información relevante, sin llenar múltiples páginas con gráficos, códigos y estadísticas.



Descripción de la Base de Datos

| Variable | Descripción |
|--------------|-----------------------------------------------------------|
| PIB | PIB per cápita - miles de millones de USD / habitante |
| POB | Población - millones |
| IDH | Índice de desarrollo humano |
| GINI | Coeficiente de Gini |
| IPC | Índice del precio al consumidor |
| FAO | Índice de precios Alimenticios de la FAO |
| GENERO | Índice de desigualdad de género |
| ELECTRICIDAD | Tasa de electrificación |
| ESCOLARIDAD | Años de escolaridad promedio |
| SUICIDIOFEM | Tasa de suicidios femeninos - cada 100.000 personas |
| SUICIDIOMAS | Tasa de suicidios masculinos - cada 100.000 personas |
| BOSQUE | Porcentaje de superficie que corresponde a bosques |
| FOSIL | Porcentaje de uso de combustibles fósiles |
| DIOXIDO | Emisiones de dióxido de carbono - Toneladas per cápita |
| DESASTRE | Población afectada por desastres naturales - miles |
| AFECTADOS | Población sin hogar por desastres naturales - miles |
| HOMICIDIO | Tasa de homicidios - cada 100.000 personas |
| MORTINF | Mortalidad de niños menores de 5 años - miles |
| MORTMAT | Tasa de mortalidad maternal - cada 100 nacimientos |
| TURISMO | Turistas internacionales - millones |
| INTERNET | Porcentaje de uso de internet |
| VIOLENCIA | Porcentaje de víctimas de violencia entre parejas |
| VIDA | Expectativa de vida - años |
| CELULAR | Subscripciones a telefonía celular - cada 100 personas |
| DESERCION | Tasa de deserción escolar primaria |
| PRISION | Tasa de encarcelamiento - 100.000 personas |
| RENOVABLE | Porcentaje de uso de energías renovables |
| PARLAMENTO | Porcentaje de escaños parlamentarios ocupados por mujeres |
| INMIGRANTES | Porcentaje de población que es inmigrante |



Pontificia Universidad Católica de Chile
Educación Profesional - Escuela de Ingeniería
Diplomado en Big Data y Ciencias de Datos
Minería de Datos
Relator: Sebastián Raveau

Su entrega será evaluada de acuerdo con los siguientes aspectos:

Claridad [1,5 puntos]

Se espera que el reporte sea claro y bien redactado, con las ideas debidamente desarrolladas. Las tablas y figuras deben estar debidamente presentadas y explicadas. El procedimiento de modelación se explica adecuadamente. El lenguaje utilizado debe ser apropiado para un reporte técnico.

Análisis Preliminar [1,5 puntos]

Se analiza y discute el impacto esperado de las potenciales variables explicativas. Se analiza y discuten potenciales transformaciones de variables explicativas (ya sean no-lineales o categóricas). Se plantean hipótesis a ser probadas en la modelación

Modelación [2 puntos]

La describe la selección de variables explicativas. Se describe el proceso de especificación del modelo.

Análisis de Resultados [1 punto]

Se discuten y analizan las hipótesis preliminares. Se analizan los resultados del modelo