

*Marco Bayesiano para el análisis de datos,
calibración de parámetros y modelamiento inverso*

Introducción

Universidad Industrial de Santander
U18 Fest

David A. Barajas-Solano

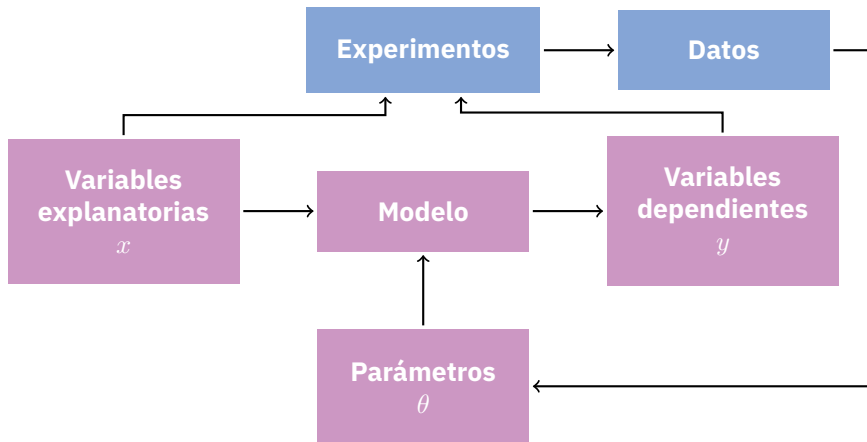
- Egresado de la UIS – Ingeniería Civil
- Ph.D. en ciencias de la ingeniería, UC San Diego
- Científico en Pacific Northwest National Laboratory
- Perfil científico en <https://www.dbarajassolano.com>

Participantes

Modelamiento en ciencia e ingeniería

- *Modelos* son objetos matemáticos usados para describir fenómenos en ciencia e ingeniería
- Los modelos son *calibrados* utilizando datos experimentales

Modelamiento en ciencia e ingeniería



Ejemplos de modelos

- *Ley de Hooke*

$$\sigma = E\varepsilon$$

- ▶ Parámetros: Módulo de Young E

Ejemplos de modelos

■ Ley de Hooke

$$\sigma = E\varepsilon$$

- ▶ Parámetros: Módulo de Young E

■ Leyes de reacción

$$R = k[A][B]$$

- ▶ Observaciones indirectas e.g. respuesta química vs tiempo
- ▶ Parámetros: Tasa de reacción k

Ejemplos de modelos

■ Ley de Hooke

$$\sigma = E\varepsilon$$

- ▶ Parámetros: Módulo de Young E

■ Leyes de reacción

$$R = k[A][B]$$

- ▶ Observaciones indirectas e.g. respuesta química vs tiempo
- ▶ Parámetros: Tasa de reacción k

■ Modelo logístico para variables categóricas

$$\log \frac{P(y = 1)}{P(y = 0)} = \alpha + \beta x$$

- ▶ Variable categórica dependiente $y = 0, 1$
- ▶ Variable explicativa continua x
- ▶ Parámetros $\theta = \alpha, \beta$

- **Modelamiento *probabilístico* para la selección, calibración, y evaluación de modelos científicos**
 - ▶ *Selección*: Cuál modelo utilizar?
 - ▶ *Calibración*: Qué valores utilizar para los parámetros?
 - ▶ *Evaluación*: Es el modelo calibrado útil?
- *Marco Bayesiano* para modelamiento probabilístico
- *Programación probabilística* para la práctica
- Algunos modelos sencillos con aplicaciones

Fundamentos

- Teoría de la probabilidad
- Teorema de Bayes

Regresión

- Modelos lineales
- Modelos no lineales
- Modelos jerárquicos

- Matemáticas y código
 - ▶ Vamos a programar? **No**
 - ▶ Vamos a calcular integrales? **No**
- Vamos a utilizar herramientas de programación probabilística para explorar conceptos

Herramientas de programación basadas en texto plano

- Lenguaje de programación Python
- Librería de programación probabilística PyMC3
- Interfase de cuadernos Jupyter Lab

Programación basada en texto plano

datos.txt

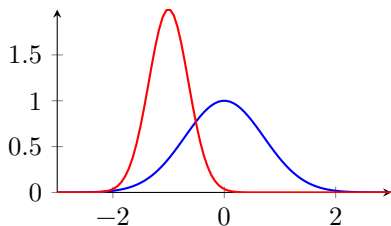
```
x1,    x2,    y
1.0,   2.6,   12.9
1.3,   8.5,   13.8
2.8,   7.1,   4.7
...
```

codigo.py

```
alpha = pm.Normal(...)
beta  = pm.Normal(...)
sigma = pm.HalfNormal(...)
mu = alpha + beta[0]...
...
```

```
python codigo.py ...
```

Interfase



Resultados

Ventajas

- *Reproducibilidad*: Se puede recrear mi análisis y resultados?
- *Transparencia*: Se puede ver cómo el análisis se ha hecho?
- *Mantenibilidad*: Se puede acceder, cambiar, o actualizar el análisis?

Alternativas

- Software de oficina

Ejemplo: Éste taller!

- *Reproducibilidad*: Software libre para generar todo el contenido (presentaciones, máquina virtual, cuadernos, etc.)
- *Transparencia*: Texto plano es transparente
- *Mantenibilidad*: Via Git

- Cronograma
 - ▶ Dos mitades de 50m–1h 20m
 - ▶ Pausa de 20m entre mitades (café y aromáticas afuera)
- Código, cuadernos y presentaciones disponibles en GitHub
<https://github.com/dbarajassolano/u18fest>
- **Preguntas bienvenidas a cualquier momento!**