

## Trabajo Final (Práctica)

*Prof: Juan Carlos Martínez Ovando**14 de noviembre de 2017*

Descarguen el archivo de datos `data_diabetes.csv` del repositorio Github del curso. Esta es la liga:

[https://github.com/jcmartinezovando/est46111\\_datos/](https://github.com/jcmartinezovando/est46111_datos/)

Los datos hacen referencia a una colección de mediciones de niveles de glucosa en la sangre en un conjunto de 70 pacientes diagnosticados con diabetes tipo II. El archivo `README.md` del mismo sitio incluye una descripción del contenido de los datos.

Estos datos los analizamos en el curso empleando el modelo base:

$$y_{it} | \alpha, \beta \sim \text{Ga}(y | \alpha, \beta) \quad (1)$$

$$\alpha, \beta \sim \text{Ga}(\alpha | a_\alpha, b_\alpha) \text{Ga}(\beta | a_\beta, b_\beta), \quad (2)$$

donde  $a$  y  $b$  son hiper-parámetros fijos y conocidos (especificados por nosotros), suponiendo que los datos  $(y_{it})_{i \geq 1, t \geq 1}$  son intercambiables.

Entorno a estos datos, existe una gran cantidad de preguntas que podemos tratar de responder. Algunos ejemplos:

1. ¿ Existe discrepancia (en media o distribucional) entre las mediciones reportadas de registros escritos en papel y las que son reportadas electrónicamente?
2. ¿ Existe discrepancia entre las mediciones de glucosa de los pacientes entre las cuatro horas del día donde se miden?
3. ¿ Existe discrepancia de las mediciones de glucosa entre pacientes?
4. ¿ Existe discrepancia entre las mediciones de glucosa asociadas con la administración de insulina con respecto a las demás?
5. ¿ Existe una discrepancia clara entre los pacientes que reportan sus niveles de glucosa después de haber ingerido más alimentos de lo usual?

Estas, entre otras preguntas, pueden responderse extendiendo el modelo (1)-(2) para considerar diferentes definiciones jerárquicas.

El trabajo final consiste en elegir dos de las cinco preguntas enlistadas anteriores para responderlas con base en:

- Definir un modelo jerárquico correspondiente a las preguntas en cuestión
- Definir informativamente los hiper-parámetros  $a$ . y  $b$ .
- Modificar el código nativo en R para aprender estadísticamente del modelo con base en los datos
- Monitorear la adecuada implementación del procedimiento de aprendizaje estadístico (i.e. tasas de convergencia y demás procedimientos de validación, como validación cruzada)
- Interpretar los resultados obtenidos después de plantear las preguntas como un problema de decisión bajo incertidumbre.

Diviértanse...