

## Clase 21 de Mayo

---

Consideremos los datos de PIMA de la librería MASS.

1. En una primera etapa exploratoria de los datos usar la función `ggpairs` para inspeccionar la relación entre las variables. ¿Qué sugiere?
2. Ajustar un modelo de regresión logística para predecir diabetes usando todas las variables. ¿Cuánto valen las estimaciones de los coeficientes? ¿Cuáles de estos coeficientes son significativos?
3. A partir del ajuste anterior construir un intervalo de confianza de nivel asintótico 0.95 para cada uno de los coeficientes del modelo ajustado.
4. Construir una *tabla de confusión* (confusion matrix o error matrix) que consiste en una tabla de doble entrada en la que cruzan la clasificación observada de los individuos y la predicha por el modelo. Computar el porcentaje de aciertos, que se conoce como la precisión de predicción.

	Predichos	
Observados	0	1
0		
1		

5. Repetir el ítem anterior sobre la muestra de testeo Pima.te. Estime el error de predicción.
6. ¿Cuál sería el valor estimado de la probabilidad de tener diabetes para una mujer de la etnia Pima que tuvo 2 embarazos y tiene una concentración de glucosa de 100, una presión sistólica de 70, un valor de piel de triceps de 20, un índice de masa corporal de 26, un valor de la función pedigree de 0.24 y 30 años de edad?
7. Si se usa el ajuste realizado: ¿cuánto vale la diferencia de log odds cuando la glucosa aumenta en 10 unidades y todos los demás valores permanecen constantes? ¿cuánto vale el cociente de los odds cuando la glucosa aumenta en 10 unidades y todos los demás valores permanecen constantes? Hallar intervalos de confianza de nivel aproximado 0.95 para el parámetro estimado en ambos casos.
8. Implementar una función que dada la salida de un ajuste del procedimiento `glm`, compute un intervalo de confianza de nivel aproximado  $1 - \alpha$  para la probabilidad de éxito en un punto  $x$ .
9. Calcular un intervalo de confianza de nivel aproximado 0.95 para la probabilidad del ítem f).