

# Examen #3

## Métodos Cuantitativos Aplicados, CIDE A.C.

### Instrucciones

El examen contiene dos (2) preguntas sobre modelos de máxima verosimilitud (80 puntos en total). Tienes 2 horas para contestarlo. Te recomiendo leer las preguntas con cuidado y tratar de no dejar respuestas en blanco. Si no tienes la respuesta final, siempre es mejor mostrar el procedimiento que tu crees que te llevaría a dicha respuesta.

Al terminar, deberás enviar el archivo con tus respuestas a mi dirección de correo electrónico ([sergio.bejar@cide.edu](mailto:sergio.bejar@cide.edu)). ¡Buena suerte!

### Pregunta 1

En esta pregunta construirás un modelo que explique el voto nominal de legisladores respecto al Tratado de Libre Comercio de Norteamérica (TLCAN o NAFTA). Para ello utilizarás la base de datos llamada `nafta.dta` que se encuentra en la carpeta Files del repositorio de Github de la clase.

La variable voto (denotada como `vote`) indica que 234 legisladores votaron a favor y 200 lo hicieron en contra. Además de dicha variable que toma valores de `yes` y `no`, la base cuenta con otras tres variables:

- `democrat`: dummy de membresía partidista (demócrata o republicano).
- `pcthispc`: porcentaje de la población que es de origen latino en el distrito de donde viene el representante.
- `cope93`: indicador COPE de posiciones pro-sindicato de cada legislador (0=anti-sindicatos, 100=pro-sindicatos).

1. Recodifica las variables (i) `vote` de tal forma que tenga el valor de 1 cuando el legislador votó a favor y 0 en cualquier otro caso y (ii) `democrat` para que tenga el valor de 1 para un legislador demócrata y 0 en cualquier otro caso. (5 pts.)

2. Estima un modelo con una función de distribución acumulada logística que te permita probar las siguientes hipótesis:

- Mayores valores COPE (más pro-sindicalismo) se asocian con una menor probabilidad de votar por NAFTA.
- El efecto de COPE sobre la probabilidad de votar por NAFTA depende de la afiliación partidista: el efecto será mayor para Demócratas que para Republicanos.

Tu modelo además deberá controlar por el porcentaje de la población que sea de origen latino en el distrito que representa el legislador. (10 pts.)

3. Presenta los resultados de tu modelo en una tabla. (5 pts.)
4. Interpreta substantivamente los resultados de forma que te permitan validar o rechazar las hipótesis nulas. (10 pts.)
5. Presenta resultados que evalúen las hipótesis utilizando gráficas de probabilidades esperadas/predicciones. (15 pts.)

## Pregunta 2

Nuestro objetivo en esta pregunta es modelar la frecuencia con la que un hilo se rompe durante el tejido. Para ello debes cargar la librería `datasets` y llamar los datos de la siguiente manera:

```
library(datasets)
data <- warpbreaks
```

La descripción de las variables es la siguiente:

- **breaks**: es una variable numérica que denota el número de roturas.
  - **wool**: es un factor que denota el tipo de lana (A o B).
  - **tension**: es un factor que denota el nivel de tensión (L, M, H). L, M, H significan: low (bajo), medium (medio), high (alto).
1. Genera un histograma de la variable **breaks** y explica que tipo de distribución observas en ella. (5 pts.)
  2. Estima la media y la varianza de la variable **breaks**. (5 pts.)
  3. Calcula un modelo que permita predecir la frecuencia con la que se rompe un hilo como función del tipo de lana y el nivel de tensión. (10 pts.)
  4. Explica detalladamente por qué decidiste usar ese modelo en lugar de otros (i.e. por ejemplo un modelo BN en lugar de OLS ó un modelo Poisson en lugar de un BN). (10 pts.)

5. Interpreta substantivamente el (los) coeficientes de tensión. (10 pts.)