

# Tarea IX

Por entregar el martes 21 de noviembre.

1. Considera el siguiente conjunto de datos:

Description:

The data were collected by the Economic Research Department of the Union Bank of Switzerland. They represent the economic conditions in 48 cities around in world in 1991.

Number of cases: 48

Variable Names:

1. City: City name
2. Work: Weighted average of the number of working hours in 12 occupations
3. Price: Index of the cost 112 goods and services excluding rent (Zurich = 100)
4. Salary: Index of hourly earnings in 12 occupations after deductions (Zurich = 100)

¿Se pueden identificar grupos en estos datos? Usa todas las herramientas que vimos en la clase. El resultado debe ser un reporte de 2 páginas.

2. Considera el siguiente método para estimar el tamaño ( $N$ ) de una población de animales de un especie particular. Primero se capturan  $M$  animales, los marcan y son puestos de nuevo en libertad.

Un tiempo más tarde se capturan animales hasta encontrar un animal marcado. Sea  $X$  el número total de animales capturados (es decir  $X$  incluye el animal marcado). Después se dejan todos los animales en libertad. Se repite lo anterior de tal forma que se obtenga una muestra  $\{x_1, \dots, x_n\}$  de  $X$  (así este procedimiento puede tardar bastante).

Puedes suponer que en cada momento la probabilidad de capturar un animal marcado es siempre igual (así se supone que  $N$  es mucho mayor que  $M$ ).

- Demuestra que

$$P(X = x) = \frac{M}{N} \left(1 - \frac{M}{N}\right)^{x-1}, x = 1, 2, \dots$$

- Demuestra que

$$\hat{\Theta}_n = \frac{M}{n} \sum_{i=1}^n X_i,$$

es el estimador de Máximo Verosimilitud. ¿Está insesgado? ¿Qué puedes decir si  $n \rightarrow \infty$ ?

- Verifica con simulaciones la distribución de  $\hat{\Theta}_n$ .
3. Se tiene interés en estimar el peso promedio de un paquete de *Principes*. Basado en el peso de una muestra de 3 paquetes,  $X_1, X_2, X_3$  se definen los siguientes dos estimadores para el peso promedio:

$$\Theta_1 = \frac{X_1 + X_2 + X_3}{3}$$

y

$$\Theta_2 = \frac{X_1 + 2X_2 + X_3}{4}.$$

- Suponiendo que  $\text{var}(X) < \infty$ , ¿Cuál de los dos estimadores prefieres y por qué?
  - Supongamos que para una muestra particular, se obtiene que  $\Theta_1 = 0,5$  y  $\Theta_2 = 0,4$ . ¿Se puede decir cual estimación será *mejor*? Explícate.
4. Un método muy antiguo de criptografía es el de la substitución: cada letra en el texto por transmitir se cambia por otra según una tabla dada.

Nos limitamos a las vocales.

En inglés se sabe que la probabilidad de que una letra arbitraria de un texto sea un **a** es 8%, un **e** es 13%, un **i** es 7%, un **o** es 8% y un **u** es 3%.

- Calcula la distribución de una letra condicional al hecho de que es una vocal.
- Interceptas un texto y obtienes que de las 100 vocales 20% es un **a**, 16% un **e**, 10% un **i**, 15% un **o**, 39% es un **u**.  
Usa el método de máxima verosimilitud para descifrar el código (suponiendo independencia).

Supongamos que parte del texto es:

.o.u.xo.x.aax..a..x.ax.a...eo.

(donde . refiere a un consonante y x a un espacio)

?Qué harías?