

Análisis exploratorio de datos

2º cuatrimestre 2013

## Trabajo nº 2

Los datos del archivo *IngresosHogar.xls* corresponden a la Encuesta Permanente de Hogares (EPH) del INDEC, 4º trimestre de 2012 (*Para más información: <http://www.indec.gov.ar> -> Trabajo e Ingresos -> Bases de Microdatos de la EPH*).

En el archivo figuran todos los hogares encuestados para los cuales se indica:

- el aglomerado urbano (AGLOMERADO);
- la cantidad de miembros del hogar (MIEMBROS);
- el ingreso total familiar (ITF).

Una de las variables directamente relacionadas con el nivel de vida de las personas de un hogar es el ingreso per cápita (ingreso total familiar / cantidad de miembros del hogar). Por este motivo, interesa comparar los ingresos per capita de los hogares de los diferentes aglomerados urbanos incluidos en la EPH.

- a) Obtenga box-plots paralelos por cada aglomerado para el ingreso per capita familiar, ordenados de acuerdo a la mediana de cada aglomerado. (Ver codificación de aglomerados al final.)
- b) Construya un gráfico de dispersión nivel.
- c) Elija la/s potencia/s que resultan de ajustar una recta por cuadrados mínimos en el gráfico de dispersión nivel. Compare la pendiente obtenida con una trazada a mano.
- d) Realice las transformaciones que sugiere el gráfico de dispersión nivel.
- e) Rehaga el punto a) para cada una de las transformaciones anteriores y elija la que considere la más adecuada. Justifique su elección brevemente.

**Sugerencia:** tenga en cuenta la función **split**; recuerde que **boxplot** y **sapply** pueden recibir una lista como argumento. Use el "help"

## **Codificación de Aglomerados**

2 = Gran La Plata  
3 = Bahía Blanca - Cerri  
4 = Gran Rosario  
5 = Gran Santa Fé  
6 = Gran Paraná  
7 = Posadas  
8 = Gran Resistencia  
9 = Cdno. Rivadavia – Rada Tilly  
10 = Gran Mendoza  
12 = Corrientes  
13 = Gran Córdoba  
14 = Concordia  
15 = Formosa  
17 = Neuquén – Plottier  
18 = S.del Estero - La Banda  
19 = Jujuy - Palpalá  
20 = Río Gallegos  
22 = Gran Catamarca  
23 = Salta  
25 = La Rioja  
26 = San Luis - El Chorrillo  
27 = Gran San Juan  
29 = Gran Tucumán - T. Viejo  
30 = Santa Rosa - Toay  
31 = Ushuaia - Río Grande  
32 = Ciudad de Buenos Aires  
33 = Partidos del GBA  
34 = Mar del Plata - Batán  
36 = Río Cuarto  
38 = San Nicolás – Villa Constitución  
91 = Rawson – Trelew  
93 = Viedma – Carmen de Patagones