

## Examen Final de Estadística Computarizada

Indicaciones:

- Los archivos requeridos son aquellos de extensión .smcl y .do (es decir los archivos log y do que se generen). No se necesita la base de datos.
- En estos archivos se debe desarrollar los comentarios y análisis requeridos.

### TEMA 1

La base “Datos.xls” contiene un conjunto de datos sobre contribuyentes del Servicio de Rentas Internas. Está en formato Excel. La segunda pestaña muestra las variables y una breve descripción. Asuma normalidad en las variables de la base de datos.

- Importe la base de datos al Stata. Agregue las etiquetas pertinentes
- Codifique las variables referentes al sector económico, al tipo de sociedad, a las provincias y al hecho de que posea vehículo o no.
- ¿Los contribuyentes de la provincia de Guayas tienen mayores ventas que los de la provincia de Pichincha?
- ¿Los contribuyentes de la provincia de Guayas, que se dedican al comercio tienen mayores ventas que los de la provincia de Pichincha, que tienen la misma actividad económica?
- ¿La proporción de sociedades que tiene vehículos registrados es mayor en las sin fines de lucro que en los contribuyentes especiales?
- ¿Los ingresos de los Grandes Contribuyentes que se dedican a la manufactura han fluctuado más que los ingresos que reciben los especiales del mismo sector?
- ¿El promedio de ingresos de las sociedades bajo el control de las superintendencias es mayor que su nivel de gastos?
- En promedio, ¿Qué grupo presenta el mayor número de oficios de sanción?
- ¿Qué grupo tiene la mayor proporción de contribuyentes, cuyo impuesto a la renta causado es igual a cero? ¿Las bajo control de las superintendencias o las otras con fines de lucro?
- ¿Qué variable presenta mayor dispersión? Las ventas o las compras de un gran contribuyente

### TEMA 2

La base “housing.dta” contiene un conjunto de datos sobre el precio de casas vendidas en los Estados Unidos. Resuelve las siguientes interrogantes.

- Se dice que en un estudio determinaron que el precio está explicado por:

$$\text{price\_sold} = \beta_0 + \beta_1 \text{sqft\_f} + \beta_2 \text{beds} + \beta_3 \text{age} + \mu$$

¿Tiene lógica los signos de los coeficientes? Como se interpretaría cada uno de ellos? ¿Son significativos los resultados? ¿Cómo se interpretaría el R cuadrado del modelo?

- En algunos casos, se puede pensar que el precio de venta de una casa está relacionado inversamente con la edad de la casa. Es decir casas que no son nuevas se venden en un valor marginalmente decreciente. Sin embargo, las casas que son muy antiguas más bien tienden a mostrar precios muy altos de venta. Se puede pensar que existe una relación cuadrática entre los precios y la edad. Pruebe la siguiente relación:

$$\text{price\_sold} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \mu$$

¿Es significativo el coeficiente de la variable elevada al cuadrado?

- c) Si intuitivamente se dice que el precio de venta de las casas puede estar explicado por los pies cuadrados, el ingreso de los habitantes del sector y el tipo de casa (finca o no), ¿serían significativas las variables incluidas?
- d) Construya un modelo donde el logaritmo de los años expliquen al logaritmo del precio de venta. El coeficiente de la pendiente, se lo considera una elasticidad. ¿En qué porcentaje aumenta el precio de las casas cuando aumenta en 1% los años de la casa?