

Examen de Mejoramiento de Estadística Computarizada

Nombre: _____ Fecha: 18/02/2011

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. (64 puntos)

Configuraciones iniciales: Grabe el archivo en formato csv e impórtelo con la aplicación Stata. Utilice como delimitador de columnas, en Stata el carácter ";". Codifique las variables referentes al sector económico, al tipo de sociedad y a la provincia.

- a) ¿Los contribuyentes de la provincia de Guayas tienen mayores ventas que los de la provincia de Pichincha?
Anote el P-value de la hipótesis alternativa que la diferencia es distinta de cero _____
Anote su conclusión: _____
- b) ¿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?
Anote el P-value de la hipótesis alternativa que la diferencia es distinta de cero _____
Anote su conclusión: _____
- c) ¿La proporción de sociedades que tiene vehículos registrados es mayor en las sin fines de lucro que en los contribuyentes especiales?
Anote el intervalo de confianza de la diferencia entre las proporciones _____
Anote su conclusión: _____
- d) ¿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?
Anote el P-Value de la hipótesis alternativa que el ratio es mayor que 1 _____
Anote su conclusión: _____
- e) ¿El promedio de ingresos de las sociedades bajo el control de las superintendencias es mayor que su nivel de gastos?
Anote el intervalo de confianza de la diferencia entre las variables _____
Anote su conclusión: _____
- f) En promedio, ¿Qué grupo presenta el mayor número de oficios de sanción?
Grupo: _____
Promedio: _____
- g) ¿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?
Anote el intervalo de confianza de la diferencia entre los grupos _____
Anote su conclusión: _____
- h) ¿Qué variable presenta mayor dispersión? Las ventas o las compras de un gran contribuyente
Anote el P-Value de la hipótesis alternativa que el ratio es mayor que 1 _____
Anote su conclusión: _____

TEMA 2.- La base “housing.dta” contiene un conjunto de datos sobre el precio de casas vendidas en los Estados Unidos. Resuelva las siguientes interrogantes. (36 puntos)

a) Se determinó que el precio está explicado por: $\text{price_sold} = \beta_0 + \beta_1 \text{sqft_f} + \beta_2 \text{beds} + \beta_3 \text{age} + \mu$. ¿Cuál es el valor del R cuadrado de este modelo? Como se interpretaría? ¿Qué variable no resulta significativa? ¿Por qué?

b) 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: $\text{price_sold} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \mu$ ¿Cuál es valor del coeficiente que acompaña a la variable elevada al cuadrado? ¿Es significativa? ¿Por qué?

c) Si intuitivamente se dice que el precio de venta de las casas puede estar explicado por los pies cuadrados y si tiene piscina o no (pool), ¿Resulta significativa la variable de la presencia o no de piscina? ¿Por qué?

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 varia el precio de las casas cuando aumenta en 1% los años de la casa? ¿Es significativa? ¿Por qué?
