

Conceptos básicos de inferencia estadística | Semestre 2018-1

Tarea 1

Fecha de entrega: 24 de agosto

1. Sean $x_1, \dots, x_n, y_1, \dots, y_n$ observaciones. Mostrar las siguientes igualdades

$$a) \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \sum_{i=1}^n x_i^2 - n\bar{x}_n^2,$$

$$b) \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) = \sum_{i=1}^n x_i y_i - n\bar{x}_n \bar{y}_n.$$

2. En el ejercicio anterior, suponer que se definen $z_i = a + bx_i$, para $i = 1, \dots, n$, con a y b constantes. Mostrar las siguientes igualdades

$$a) \bar{z}_n = a + b\bar{x}_n,$$

$$b) s_z^2 = b^2 s_x^2.$$

$$\text{donde } \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i, s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \text{ y } s_z^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z}_n)^2.$$

3. Del archivo de descripción de las bases de microdatos de la Encuesta Intercensal 2015 del INEGI (disponible [aquí](#)), seleccionar 9 variables de la base de datos de viviendas (hoja TR_Viviendas) y 9 variables de la base de datos de personas (hoja TR_personas), de manera que se incluyan: 3 nominales, 3 ordinales y 3 en escala de intervalo o de razón. **No considerar las variables de identificación (llaves, claves geográficas, identificadores, variables del diseño de muestreo).**

4. El archivo muestraENIGH.csv (disponible en [aquí](#)) contiene una muestra de una de las bases de microdatos de la Encuesta Nacional de Ingresos y Gastos en los Hogares (ENIGH) realizada por el INEGI en 2014. Cada registro contiene la información de un hogar.

Utilizar la información en el archivo anterior para responder las siguientes preguntas. Para como referencia el documento *Descripción de la bases de datos* del INEGI (disponible ([disponible aquí](#)) la tabla a consultar es CONCENTRADOHOGAR.

a) ¿Cuántas variables y observaciones contiene el archivo?

b) Calcular el promedio y la mediana del número de integrantes del hogar.

c) Calcular el promedio y la mediana de las edades de los jefes del hogar, ¿son iguales?

d) Graficar un histograma y la densidad estimada de las edades de los jefes del hogar.

e) Calcular la proporción de jefes del hogar con y sin **educación básica completa**, ¿qué proporción es mayor?

f) Calcular el promedio del ingreso corriente en los hogares según la escolaridad del jefe del hogar, considerar únicamente las categorías educación básica incompleta y educación básica completa, ¿hay alguna diferencia?

g) Graficar las densidades estimadas del ingreso corriente en los hogares según la escolaridad del jefe del hogar. De nuevo, considerar únicamente las categorías educación básica incompleta y educación básica completa. ¿En que categoría hay mayor variabilidad en los ingresos?