

Project 3

May 28, 2019

Proyecto 3

Ingeniería en Estadística, Universidad de Valparaíso

Profesor: Eduardo Jorquera - eduardo.jorquera@postgrado.uv.cl

Fecha y hora de entrega: 3 de Junio del 2019, 23:55 hrs

La penalización por entregarla tarde es de un punto menos de la nota final

La tarea debe ser entregada en formato .html y .ipynb (ir a File -> Download as)

1 Instrucciones

1.- Genere un dataframe de pandas donde la primera variable sea la id de la observación, la segunda variable sea generada desde una distribución normal estándar, de tamaño 5.000. Haga un histograma utilizando y guárdelo en un archivo separado por comas (.csv). Muestre un histograma calculando el número de barras usando la fórmula de Sturges.

2.- Al dataframe anterior, calcule la media, desviación estándar, varianza, moda y mediana.

3.- Genere un nuevo dataframe de pandas, el cual contenga 5 grupos de 20 observaciones. Genere el primer grupo desde una distribución normal con media 0 y varianza 2; el segundo grupo desde una distribución normal con media 1 y varianza 1; el tercer grupo desde una distribución Uniforme(-10,10); el cuarto y quinto grupo desde una distribución normal con media 4 y 5, y varianza 2 y 3, respectivamente. Para esto tendrá que indexar las observaciones y los grupos.

4.- Haga un histograma por cada grupo, y calcule los mismos estadísticos mencionados en 2.-.

5.-

- Simule el peso de una población desde una distribución normal con media 70 kg y varianza 30 y guárdelo en una serie de pandas de largo de 500 observaciones.
- Luego cree otra serie de pandas que indicará si la persona tiene la educación media completa. Esta variable será 0 ó 1 desde una distribución Uniforme(0,1), donde si es mayor a 0.5, entonces el valor en la variable será 1; y si es menor o igual a 0.5, entonces en la variable toma el valor 0.
- Junte estos objetos en un solo dataframe, con una nueva variable que será la indexación de los datos, y luego haga un gráfico de torta y de barra para mostrar la proporción de unos y ceros (1 y 0) de este conjunto de datos.
- Grafique el peso de ambas poblaciones a través de dos histogramas.

6.-

- Cree un nuevo dataframe del mismo largo que el anterior que contenga dos variables: la indexación y la altura.

- Genere la altura a partir de una distribución normal con media 1.7 m y varianza 0.4.
- Junte este nuevo dataframe con el dataframe resultante de 5.- respecto a la variable de indexación.