

Final Project

June 18, 2019

Proyecto final

Ingeniería en Estadística, Universidad de Valparaíso

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

Fecha y hora de entrega: 24 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

Cree un menú que permita elegir una de las siguientes opciones:

1.- Cree una función que retorne la media y la varianza sin hacer uso de alguna librería o paquete. Luego, use esta función en un arreglo de numpy creado por el usuario cuyas dimensiones son especificadas por él, y calcule la media y la varianza usando la función creada según la primera fila y la primera columna del arreglo.

2.- Cree una función que permita al usuario ingresar n valores enteros en el intervalo $[a, b]$, ó generarlos aleatoriamente desde una distribución uniforme con los mismos límites, donde a, b y n son ingresados por el usuario y n debe ser mayor o igual a 20. Calcule la media y la varianza de estos datos utilizando la función ya creada en 1.-.

3.- Construya una dataframe de pandas donde estén los datos generados en 2.-. Agregue una columna que indique si la observación está por sobre o por debajo del promedio (con 1 y 0), cree un gráfico de torta de esta nueva variable. Guarde estos datos en formato .csv.

4.- Haga que el usuario ingrese los valores de μ y σ^2 , y grafique la función

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad (1)$$

para los valores de $x \in [\mu - 3\sigma, \mu + 3\sigma]$. Guarde el archivo en formato .png.

5.- Para este problema, haga que el usuario también elija si desea ingresar $N, M, n, m \in \mathbb{N}$ o que sean generados desde una distribución uniforme. Cree una matriz con numpy de $N \times M$, donde los datos son ingresados por el usuario ó generados de una distribución normal estándar, según lo indique el usuario. Luego, dígame al usuario que saque una partición $n \times m$ de esta matriz, donde $n < N$ y $m < M$. Calcule la media por columna de esta nueva partición través de un ciclo for, y luego transforme este arreglo a un vector fila para finalmente calcular la media total de los datos.

6.- Muestre cómo obtener la frecuencia de cada valor del dataframe completo, usando:

```
In [1]: import pandas as pd; import numpy as np
df = pd.DataFrame(np.random.randint(1, 10, 20).reshape(-1, 4), columns = list('abcd'))
df
```

```
Out[1]:
```

	a	b	c	d
0	7	8	7	5
1	2	5	1	4
2	8	9	4	1
3	8	4	6	9
4	8	6	7	7