

Inicio ► Instituto del Cálculo ► 2024 ► Primer Cuatrimestre 2024 ► Labo Datos IC 24 ► Parcial ►  
Simulacro de parcial ► Vista pre...

Comenzado el Tuesday, 11 de June de 2024, 10:55

Estado Finalizado

Finalizado en Tuesday, 11 de June de 2024, 10:57

Tiempo empleado 1 min 56 segundos

Pregunta 1

Correcta

Puntúa como 1.00

¿Cuál de los siguientes comandos de Python calcula correctamente el error cuadrático medio entre el vector respuesta y el vector de predicciones?

Seleccione una:

- ☒ a. `np.sum((y - y_pred)**2) / len(y)` ✓
- ☐ b. `np.sum((y - y_pred))**2 / len(y)`
- ☐ c. `(y.mean())**2 - (y_pred.mean())**2`
- ☐ d. `np.mean(pd.concat([y, y_pred])**2)`

Pregunta 2

Correcta

Puntúa como 1.00

Ambos métodos de clustering  $k$ -medias y DBSCAN son sensibles a la escala de los datos.

Seleccione una:

- ☒ Verdadero ✓
- ☐ Falso

Pregunta 3

Correcta

Puntúa como 1.00

Clustering es un problema de aprendizaje supervisado.

Seleccione una:

- ☐ Verdadero
- ☒ Falso ✓

Pregunta 4

Finalizado

Puntúa como 1.00

Utilizando el dataset penguins, calcular:

1. Peso promedio de los pingüinos de la especie Adelie.
2. Peso promedio de los pingüinos de la especie Adelie que viven en la Isla Dream.
3. Para cada isla, el desvío estándar del peso de los pingüinos que viven en esa isla.

1. Respuesta: 3706.16
2. Respuesta: 3701.36
3. Respuesta: Biscoe: 790.86, Dream: 412.93, Torgersen: 451.84

.....

Pregunta 5

Correcta

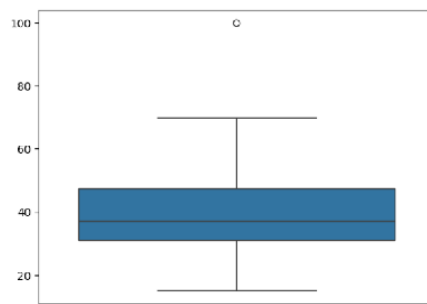
Puntúa como 1.00

Utilizando las edades de los pacientes que recibió un médico un día, se realiza un boxplot para resumir los datos (utilizando bigotes de longitud hasta 1.5 veces la distancia intercuartil).

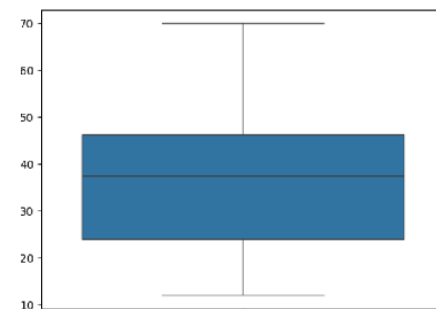
Si los datos tienen las siguientes propiedades:

- Media: 43
- Mediana: 37
- Máximo: 100
- Mínimo: 15

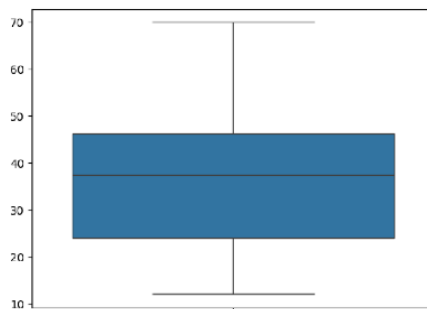
¿Cuál de los cuatro box-plot es el correcto?



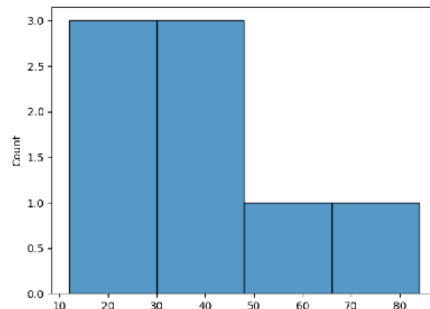
(a)



(b)



(c)



(d)

Seleccione una:

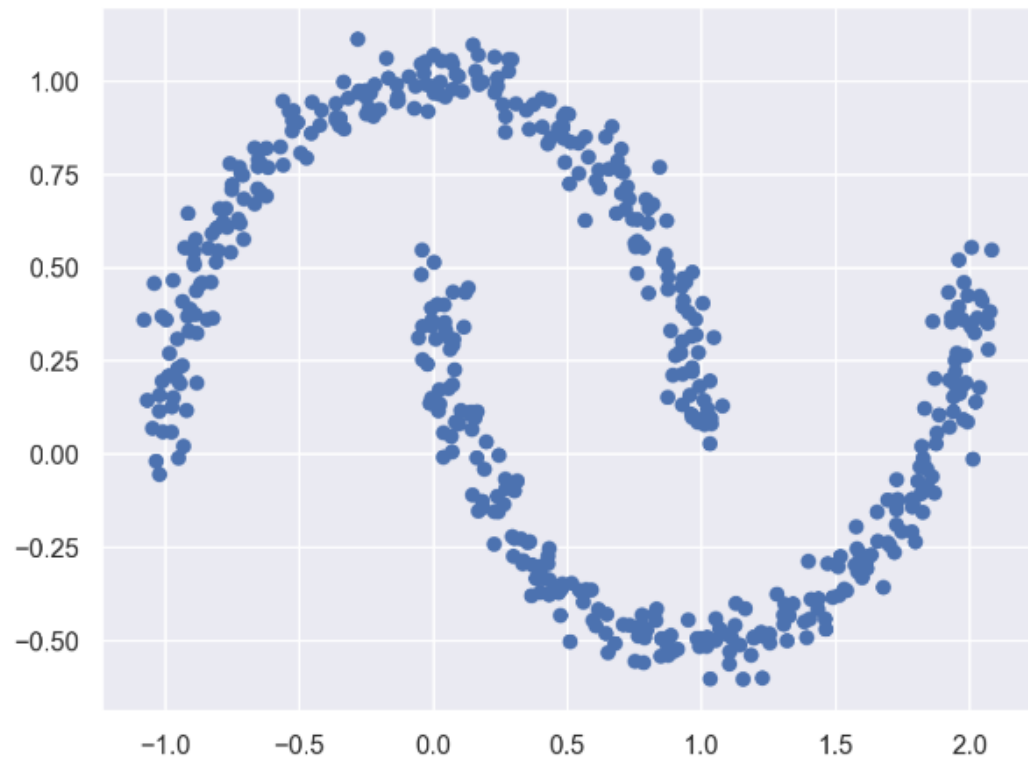
- ☒ A) ✓
- ☐ B)
- ☐ C)
- ☐ D)

Pregunta 6

Correcta

Puntúa como 1.00

Se quiere realizar un clustering con estos datos. ¿Cuál de las siguientes opciones considera que es la más apropiada?



Seleccione una:

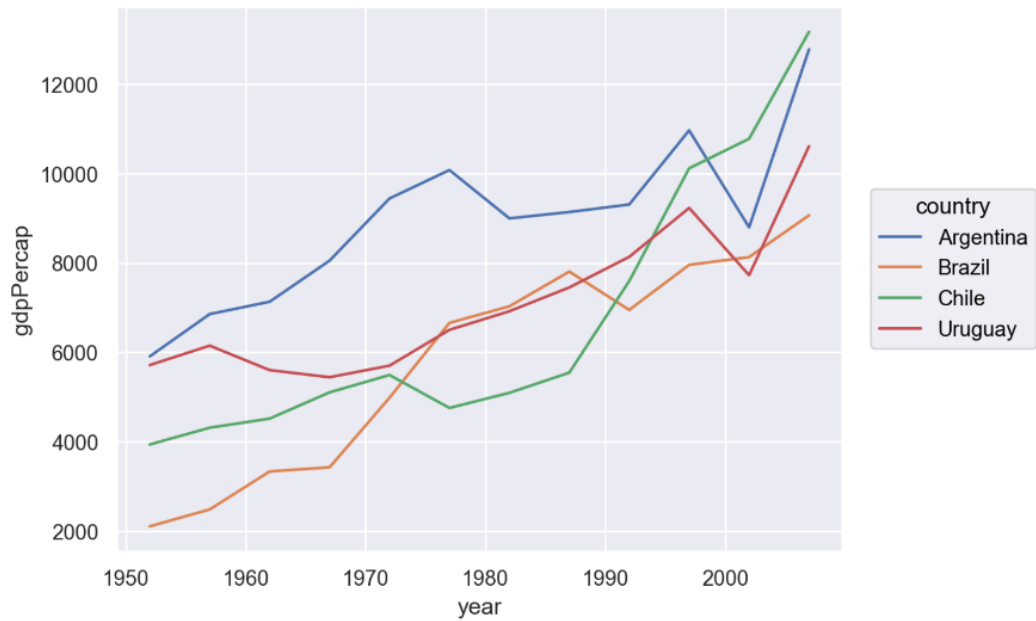
- ☐ a. DBSCAN con  $\epsilon = 0.8$  y minPts = 20
- ☐ b. K-medias con  $K = 3$
- ☐ c. K-medias con  $K = 2$
- ☒ d. DBSCAN con  $\epsilon = 0.2$  y minPts = 20 ✓

Pregunta 7

Finalizado

Puntúa como 1.00

Ingresar un código para generar el siguiente gráfico a partir de datos de Gapminder.



Pueden utilizar el siguiente código para cargar los datos de Gapminder.

```
from gapminder import gapminder
```

```
sudamerica = gapminder[gapminder['country'].isin(['Argentina', 'Brazil', 'Chile',  
'Uruguay'])]
```

```
(  
so.Plot(data=sudamerica, x='year', y='gdpPercap', color='country')  
.add(so.Line())  
)
```

```
.....
```

## Pregunta 8

Correcta

Puntúa como 1.00

La formula de Wilkinson

$$\text{body\_mass\_g} \sim (\text{bill\_length\_mm} * \text{flipper\_length\_mm}) + \text{bill\_depth\_mm}$$

representa un modelo que...

Seleccione una:

- ☐ a. Contempla la interaccion entre las variables bill\_length\_mm y bill\_depth\_mm
- ☒ b. Contempla la interaccion entre las variables bill\_length\_mm y flipper\_length\_mm ✓
- ☐ c. No tiene intercept
- ☐ d. Contempla la interaccion entre las variables bill\_depth\_mm y flipper\_length\_mm
- ☐ e. Ninguna opcion es correcta
- ☐ f. Todas las opciones son correctas

## Pregunta 9

Correcta

Puntúa como 1.00

Si se desea estimar una variable numerica (Y) por regresión lineal con intercept utilizando como variables explicativas una variable numerica (X) y una variable categorica (Z) de 3 categorias, ¿cuantas variables dummy son necesarias? (responda solo con el número)

Respuesta: 2



## Pregunta 10

Correcta

Puntúa como 1.00

Una cierta matriz de covarianza tiene los siguientes autovalores:

$\lambda_1$	15
$\lambda_2$	10
$\lambda_3$	9
$\lambda_4$	8
$\lambda_5$	1

¿Cual es la menor cantidad de componentes principales necesaria para explicar el 70% de la varianza de los datos?

Seleccione una:

- ☐ 1
- ☐ 2
- ☒ 3 ✓
- ☐ 4
- ☐ 5