



Examen Práctico

INSTRUCCIONES

- Utiliza tu computadora con Spyder para resolver los ejercicios.
- Puedes utilizar todo el material de consulta que necesites, apuntes, libros, hojas, códigos de ejercicios anteriores, etc.
- No está permitido utilizar la conexión a internet con ningún dispositivo electrónico (reloj, laptop, celular, computadora, etc.) para consultas.
- Solo puedes utilizar los módulos Pandas, Pyplot y Math. No puedes utilizar ningún otro módulo (**import**)
- Cada ejercicio vale 1 punto.
- Guarda el código de cada ejercicio como **ejercicio_1.py, ejercicio_2.py, ejercicio_3.py, ejercicio_4.py**
- Cuando termines de resolver los problemas, comprime los cinco archivos de los ejercicios en una carpeta **.zip** o **.rar** y nombra la carpeta comprimida de la siguiente manera: **apellido1_apellido2_nombre1_nombre2_nombre3.zip**, todo en **minúsculas y sin acentos**, por ejemplo: **gomez_carranza_juan_carlos.zip**
- Sube el archivo comprimido a **Teams**.

EJERCICIOS

- El archivo adjunto salaries.csv contiene datos de 5 variables de interés para puestos de trabajo en México. Las variables son (entre paréntesis está el nombre de la variable en el archivo), puesto (offer), salario (salary), horas diarias de trabajo (hours_worked), días semanales de trabajo (days_worked), y estado (state). Escribe un programa en Python que lea el archivo con Pandas y calcule lo siguiente:
 - El resumen de los cinco números, la media, la desviación estándar y dibuje el diagrama de caja para la variable salary de forma independiente para los estados de Guanajuato y Querétaro, pero solo para las personas que trabajan entre 20 y 40 horas a la semana (todos los cálculos deben ser independiente para cada estado).
 - La correlación entre las variables salary y days_worked en general de forma independiente para estado posible (cada uno de los que aparecen en el conjunto de datos).
- Con el mismo archivo del punto anterior escribe un programa en Python que:
 - Transforme la variable salary a una variable binaria, considerando sueldos altos o bajos. Un salario bajo es aquel menor o igual 10000, el alto es mayor a 10000.
 - Transforme la variable hours_worked en una variable binaria, considerando tiempo completo o tiempo parcial, el tiempo parcial es menor o igual a 6 horas, el tiempo completo es mayor a 6 horas
 - Genere una tabla de contingencia con estas dos variables como se muestra a continuación (los números son ficticios)



		Horas de diarias de trabajo	
		Completo	Parcial
Salario	Alto	10	5
	Bajo	1	4

La cantidad en cada celda es el número de respuestas en el cruce de valores de ambas variables (en el ejemplo, hay 10 puestos de trabajo de tiempo completo y ofrecen un salario alto). La tabla se debe generar como una lista de listas, una lista para cada tipo de salario, con dos valores por lista.

d) Calcular el valor esperado para cada combinación de renglón y columna (¿cuántas ofertas de un tipo de salario son para trabajar tiempo completo o parcial?). El valor esperado se calcula como

$$E_{c,r} = \frac{(\text{suma de renglón } r)(\text{suma de columna } c)}{\text{total de observaciones}}$$

- Utilizando el mismo archivo de salarios, hacer un programa en Python que para cada puesto de trabajo tome el subconjunto de variables con salary, hours_worked y days_worked, de tal forma que cada puesto de trabajo pueda representar un punto en un espacio 3D. Después, debe calcular la distancia Euclideana entre cada par posible de puntos. La salida debe ser una lista de listas de $n \times n$ con las distancias (n = número de puestos de trabajo en la lista).

Ejemplo (ficticio):

Puestos de trabajo:

20000,10,5

5000,12,4

40000,8,7

Salida:

```
[[ 0.  30.4  21.5],
 [ 30.4  0.  10.4],
 [ 21.5 10.4  0.  ]]
```

Explicación: Hay tres puntos (tres puestos de trabajo) en el archivo, cada uno con sus valores para las tres variables. La salida es una lista de listas 3x3 (3 listas con 3 elementos cada una), formando una matriz de distancias. La distancia de un punto con sí mismo es 0, por lo tanto, la diagonal contiene solo 0s.



La distancia Euclideana se calcula como:

$$d_{p,q} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Donde p y q son dos puntos cualquiera y n es la dimensión de esos puntos, en este caso 3.

4. Utilizando el mismo archivo previo, escribe un programa en Python que encuentre la(s) moda(s) para las variables salary, hours_worked y days_worked. La moda es el valor de la lista que más se repite (si varios números se repiten el mismo número máximo de veces, todos son modas).

Ejemplo (ficticio para una lista cualquiera):

Entrada:

[1, 2, 3, 4, 5, -1, -1, -5, -2, 1, 2, 1, 2, 3, 3, -1]

Salida:

Modas: 1, 2, 3, -1

Explicación: Los valores 1, 2, 3 y -1 ocurren con la misma frecuencia máxima (3) en la lista.