

✓ EJERCICIO 1

Tres trabajadores de una cadena de montaje van a comparar su efectividad. Se recoge el tiempo que tarda cada trabajador en realizar su tarea sobre las siguientes 10 unidades que pasan por su punto de la cadena.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
workerA = np.array([4.44, 4.77, 6.56, 5.07, 5.13, 6.72, 5.46, 3.73, 4.31, 4.55])
workerB = np.array([6.22, 5.36, 5.40, 5.11, 4.44, 6.77, 5.50, 3.03, 5.70, 4.53])
workerC = np.array([3.93, 4.78, 3.97, 4.27, 4.37, 3.31, 5.84, 5.15, 3.86, 6.25])
```

Tenemos que decidir si hay algún trabajador más rápido o no. Para ello calculamos las medias del tiempo.

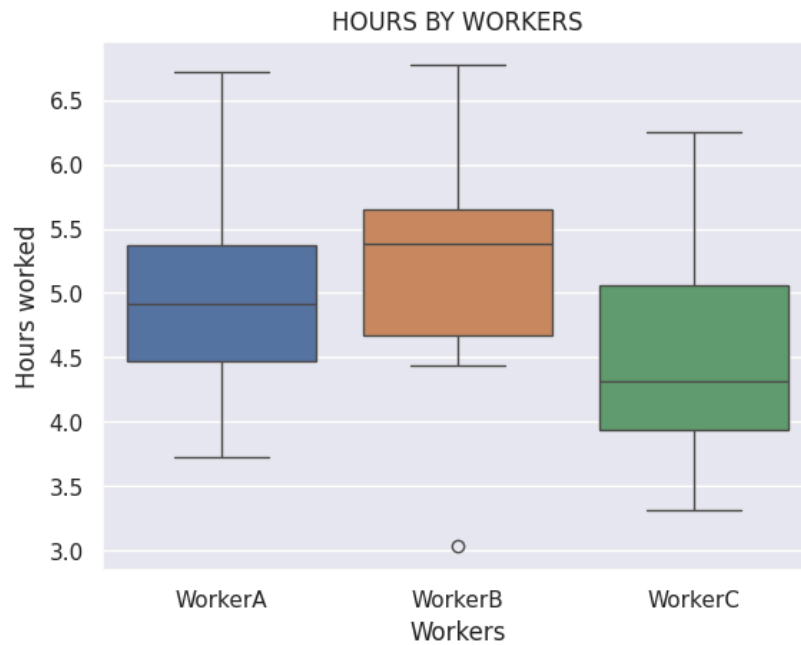
```
data = {"WorkerA": workerA, "WorkerB": workerB, "WorkerC": workerC}
df = pd.DataFrame(data)

print(
    "Mean-WorkerA: {:.4f}\nMean-WorkerB: {:.4f}\nMean-WorkerC: {:.4f}".format(
        df["WorkerA"].mean(), df["WorkerB"].mean(), df["WorkerC"].mean()
    )
)
```

```
➡ Mean-WorkerA: 5.0740
   Mean-WorkerB: 5.2060
   Mean-WorkerC: 4.5730
```

✓ Diagrama de cajas y bigotes

```
sns.set_theme(style="darkgrid")
sns.boxplot(data=df)
plt.title("HOURS BY WORKERS")
plt.xlabel("Workers")
plt.ylabel("Hours worked")
plt.show()
```



Con el "ancho" de las cajas se aprecia lo que fluctúan los tiempos de cada trabajador, y se ve que otro muestreo podría haber resultados en otro ganador.

✓ EJERCICIO 2

Se desarrollará la misma situación pero con otros datos.

```
workerA = np.array([5.01, 5.04, 5.22, 5.07, 5.08, 5.24, 5.11, 4.94, 5.00, 5.02])
workerB = np.array([5.18, 5.09, 5.09, 5.06, 5.00, 5.23, 5.10, 4.86, 5.12, 5.01])
workerC = np.array([4.51, 4.59, 4.51, 4.54, 4.55, 4.45, 4.70, 4.63, 4.50, 4.74])
```

Tenemos que decidir si hay algún trabajador "más rápido" o no. Volvemos a calcular las medias de

```
data = {"WorkerA": workerA, "WorkerB": workerB, "WorkerC": workerC}
df = pd.DataFrame(data)

print(
    "Mean-WorkerA: {:.4f}\nMean-WorkerB: {:.4f}\nMean-WorkerC: {:.4f}".format(
        df["WorkerA"].mean(), df["WorkerB"].mean(), df["WorkerC"].mean()
    )
)
```



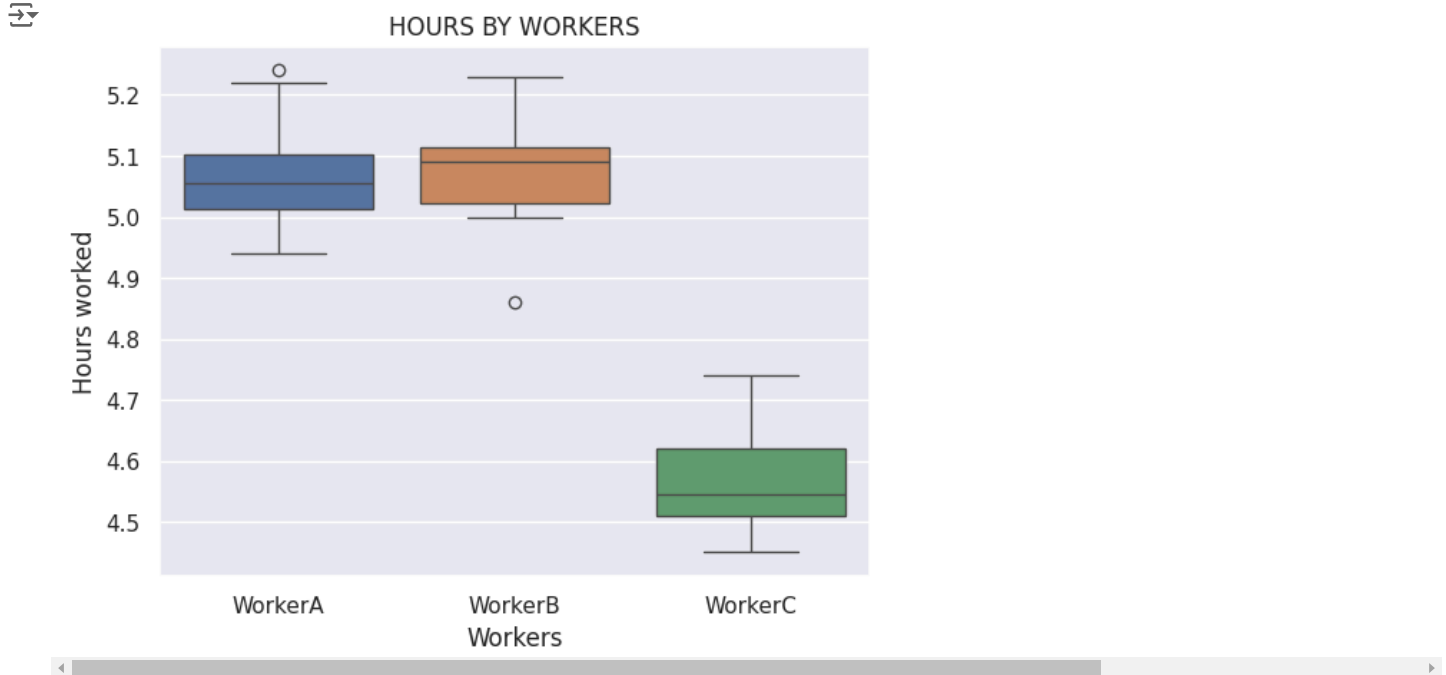
```
Mean-WorkerA: 5.0730
Mean-WorkerB: 5.0740
Mean-WorkerC: 4.5720
```

Por lo tanto, el más rápido ha sido el C.

Realizamos la configuración de boxplot. Pero se ve que los tiempos de cada trabajador fluctúan por lo que tal vez, con otra muestra de los mismos trabajadores, el ganador podría haber sido otro.

✓ Diagrama de cajas y bigotes

```
sns.boxplot(data=df)
plt.title("HOURS BY WORKERS")
plt.xlabel("Workers")
plt.ylabel("Hours worked")
plt.show()
```



Medias significativamente diferente.

✓ Ejercicio 3

```
np.random.seed(123)
n = 10
morning = np.random.normal(loc=5, scale=1, size=n)
afternoon = np.random.normal(loc=5, scale=1, size=n)
night = np.random.normal(loc=5, scale=1, size=n)
```

```
data = {"Morning": morning, "Afternoon": afternoon, "Night": night}
df = pd.DataFrame(data=data)
print(df)
```

```

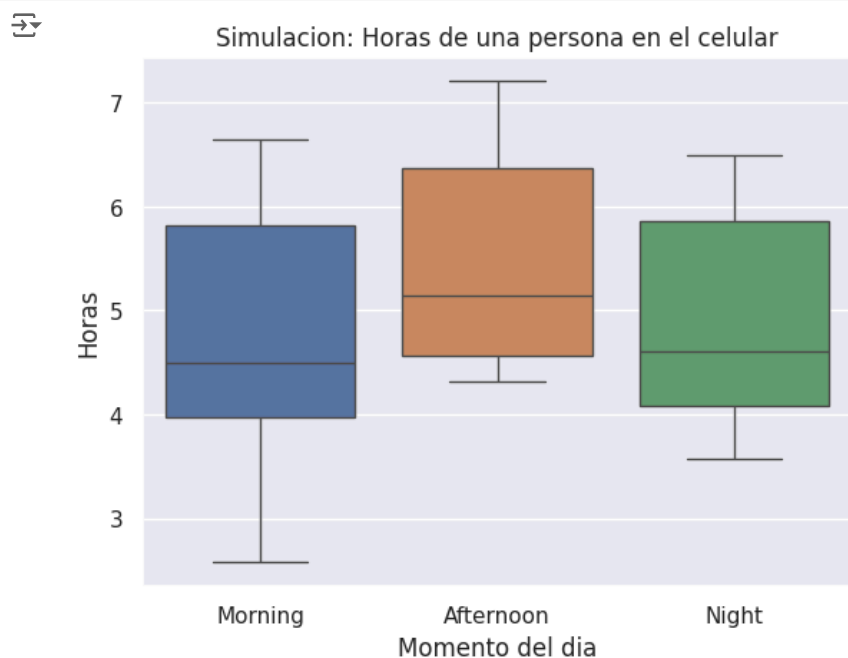
Morning  Afternoon  Night
0  3.914369  4.321114  5.737369
1  5.997345  4.905291  6.490732
2  5.282978  6.491390  4.064166
3  3.493705  4.361098  6.175829
4  4.421400  4.556018  3.746119
5  6.651437  4.565649  4.362248
6  2.573321  7.205930  5.907105
7  4.571087  7.186786  3.571319
8  6.265936  6.004054  4.859931
9  4.133260  5.386186  4.138245
```

Calculamos las medias:

```
print(
    "Mean-Morning: {:.4f}\nMean-Afternoon {:.4f}\nMean-Night {:.4f}".format(
        df["Morning"].mean(), df["Afternoon"].mean(), df["Night"].mean()
    )
)
```

↗ Mean-Morning: 4.7305
Mean-Afternoon 5.4984
Mean-Night 4.9053

```
sns.boxplot(data=df)
plt.title("Simulacion: Horas de una persona en el celular")
plt.xlabel("Momento del dia")
plt.ylabel("Horas")
plt.show()
```



En el bloque de código anterior, se ha generado 30 datos numéricos: 10 por la mañana, 10 por la tarde, 10 por la noche. Se comparan las 3 muestras visualizando los 3 diagramas de caja comparativos.

- ¿De qué distribución de probabilidad viene cada una de las 3 muestras? **DISTRIBUCION NORMAL**
- ¿Cuál es la media de cada submuestra? Mean-Morning: **4.7305** Mean-Afternoon **5.4984** Mean-Night **4.9053**
- ¿Cuál es la media de cada población? **5.0447**
- ¿Ha coincidido la media de cada muestra con la media de la población de la que proviene? **SI** ¿Cuál crees que se la razón? **PORQUE TIENEN UNA DISTRIBUCION NORMAL**
- Si te interesa conseguir valores "generalmente más altos" ¿en qué franja del día interesa muestrear? **EN LA TARDE PORQUE ESTÁ MÁS PROXIMA A LA MEDIA DE LA POBLACIÓN**

Empieza a programar o a [crear código](#) con IA.