Cosijoeza Melchor Nolasco Estadistica con Python Maestría en Ciencia de Datos Universidad Tecnológica de la Mixteca

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.graphics.factorplots import interaction_plot
```

Datos

```
\texttt{resistencia} = [15.29, 15.89, 16.02, 16.56, 15.46, 16.91, 16.99, 17.27, 16.85, 16.35, 17.23, 17.81, 17.74, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.85, 16.
                                                                                                                      18.02, 18.37, 12.07, 12.42, 12.73, 13.02, 12.05, 12.92, 13.01, 12.21, 13.49, 14.01, 13.30,
                                                                                                                      12.82,12.49,13.55,14.53]
templado = ["Rapido"] * 15 + ["Lento"] * 15
 grosor = ([8] * 5 + [16] * 5 + [24] * 5) * 2
```

Propuesta de generacion de datos aleatorios.

Templado Grosor Resistencia

```
templado = np.random.choice(a=["Rapido", "Lento"], size=30, p=[0.5, 0.5])
grosor = np.random.choice(a=[8,16,24],size=30,p=[0.35,0.35,0.3])
resistencia = np.random.normal(loc=15.3, scale=2.5, size=30)
datos = {"Templado": templado, "Grosor": grosor, "Resistencia": resistencia}
df = pd.DataFrame(data=datos)
print(df)
```

```
\overline{2}
                         18.315239
         Lento
          Lento
                          18.317865
                    16
                         15.783989
          Lento
                    24
                    16 17.015137
16 17.708248
    3
         Lento
         Lento
        Rapido
                    8 11.407043
24 15.659337
    5
    6
         Lento
        Rapido
                    16
                          16.713844
    8
        Rapido
                    24
                          15.986632
    9
         Lento
                    16
                          16.825586
    10
        Rapido
                    8
                          13.727665
    11
        Rapido
                          10.559114
    12
        Rapido
                    16
                          16.029963
    13
          Lento
                          17.854543
    14
         Lento
                     8
                          12.720293
                          13.049846
    15
        Rapido
                     8
    16
         Rapido
                    24
                          11.324937
    17
          Lento
                    24
                          14.321626
    18
         Lento
                    24
                          12.996293
    19
        Rapido
                          14.283547
    20
        Rapido
                     8
                          17.806281
    21
          Lento
                   16 17.203779
                    8
    22
          Lento
                          15.562524
    23
        Rapido
                          12.022732
    24
          Lento
                    24
                          15.339702
        Rapido
                    16
                          15.233340
                          18.300566
    26
    27
        Rapido
                          20.317422
    28
          Lento
                    16
                           17.436628
                          16.242707
```

Análisis descriptivo y gráfico

En primer lugar, se generan los diagramas Box-plot para identificar posibles diferencias notables, asimetrías, valores atípicos y homogeneidad de varianza entre los distintos niveles. Se calcula también la media y varianza de cada grupo.

```
plt.figure(figsize=(10, 6))
sns.set_theme(style="darkgrid")
```

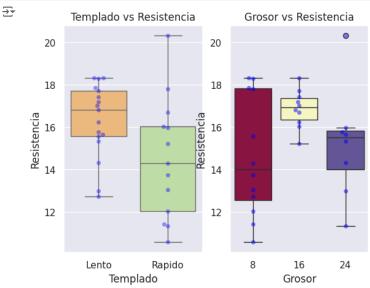
```
₹ <Figure size 1000x600 with 0 Axes>
```

Haz doble clic (o pulsa Intro) para editar

Grafica que muestra el templado con respecto a la resistencia y Grafica que muestra el grosor con respecto a la resistencia.

```
plt.subplot(1, 2, 1)
sns.boxplot(x="Templado", y="Resistencia", hue="Templado",legend=False, data=df, palette="Spectral")
sns.swarmplot(x="Templado", y="Resistencia", data=df, color="blue", alpha=0.4)
plt.title("Templado vs Resistencia")

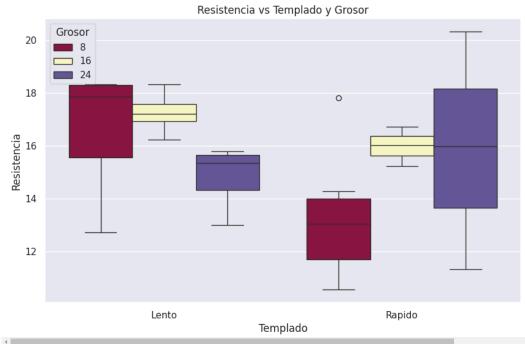
plt.subplot(1, 2, 2)
sns.boxplot(x="Grosor", y="Resistencia",hue="Grosor",legend=False,data=df, palette="Spectral")
sns.swarmplot(x="Grosor", y="Resistencia", data=df, color="blue", alpha=0.5)
plt.title("Grosor vs Resistencia")
```



Grafica que muestra la resistencia con respecto al templado y grosor.

```
plt.subplots(1, 1, figsize=(10, 6))
sns.boxplot(x="Templado", y="Resistencia", hue="Grosor", data=df,palette="Spectral")
plt.title("Resistencia vs Templado y Grosor")
```

Text(0.5, 1.0, 'Resistencia vs Templado y Grosor')



Calculo de la media y la desviacion estandar de la resistencia por cada tipo de templado.

print("Resistencia media y desviacion tipica por templado") df.groupby("Templado")["Resistencia"].agg(["mean", "std"]) Resistencia media y desviacion tipica por templado mean std \blacksquare **Templado** īl. 16.329651 1.753661 Lento Rapido 14.497105 2.869737

Calculo de la media y la desviacion estandar de la resistencia por cada tipo de grosor en cada tipo de templado.

```
print("Resistencia media y desviacion tipica por el templado y grosor.")
df.groupby(["Templado", "Grosor"])["Resistencia"].agg(["mean", "std"])
```

Resistencia media y desviacion tipica por el templado y grosor.

std

mean

Templado	Grosor		
Lento	8	16.550633	2.425085
	16	17.249993	0.662976
	24	14.820190	1.170271
Rapido	8	13.265175	2.388796
	16	15.992382	0.740967
	24	15.876330	4.497257
4			

A partir de la representación gráfica y el cálculo de las medias se puede intuir que existe una diferencia en la resistencia alcanzada dependiendo del tipo de templado. La resistencia parece incrementarse a medida que aumenta el grosor de la lámina, si bien no está claro que la diferencia en las medias sea significativa. La distribución de las observaciones de cada nivel parece simétrica sin presencia de valores atípicos. A priori, parece que se satisfacen las condiciones necesarias para un ANOVA.

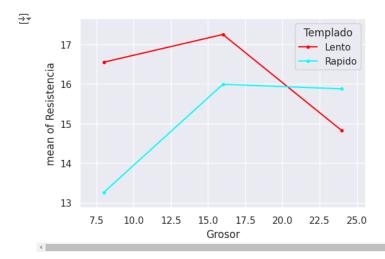
Otra forma de identificar de forma gráfica posibles interacciones entre los dos factores es mediante lo que se conocen como "gráficos de interacción". Si las líneas que describen los datos para cada uno de los niveles son paralelas, significa que el comportamiento es similar independientemente del nivel del factor, es decir, no hay interacción.

Graficos de interaccion.

```
fig, ax = plt.subplots(figsize=(6, 4))
fig = interaction_plot(x=df.Templado, trace=df.Grosor, response=df.Resistencia, ax=ax)
```

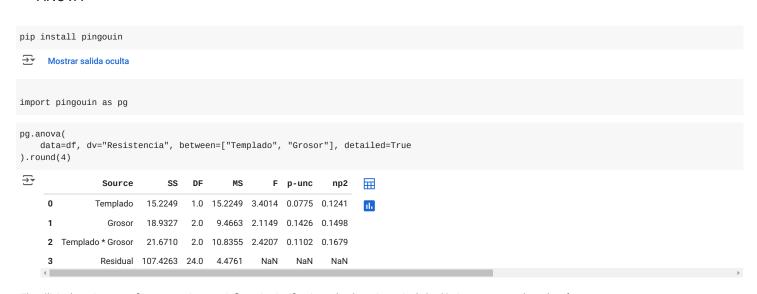


```
fig, ax = plt.subplots(figsize=(6, 4))
fig = interaction_plot(x=df.Grosor, trace=df.Templado, response=df.Resistencia, ax=ax)
```



El primer gráfico de interacción parece indicar que el incremento de resistencia entre los dos tipos de templado es proporcional para los tres grosores. En el segundo gráfico, se observa cierta desviación en el grosor de 24mm. Esto podría deberse a simple variabilidad o porque existe interacción entre las variables grosor y templado. Estos indicios serán confirmados o descartados mediante el ANOVA.

ANOVA



El análisis de varianza confirma que existe una influencia significativa sobre la resistencia de las láminas por parte de ambos factores (templado y grosor) con tamaños de efecto η2 grande y mediano respectivamente. Sin embargo, no se detecta una interacción significativa entre ellos.