

## Ejercicio 1

Una de las enfermedades que más preocupa a las autoridades sanitarias es la hipertensión. La hipertensión arterial afecta más del 30% de la población adulta mundial y es el principal factor de riesgo de padecer enfermedades cardiovasculares. Por ello, la lucha contra esta enfermedad es objetivo de los sistemas sanitarios de todo el mundo. En un laboratorio europeo se está trabajando en un medicamento innovador con cierta composición de etanol para reducir la presión arterial alta (sistólica). El medicamento se suministra a 100 pacientes de diferentes edades por vía oral en forma de comprimido. Para obtener mejores conclusiones en el estudio, se divide a las 100 personas en dos grupos, grupo 1 (pacientes con edad igual o inferior a 35 años) y grupo 2 (pacientes con edad superior a 35 años). En la hoja de cálculo Excel llamada "HIPERTENSIÓN" aparecen los datos de las 100 personas que se han empleado en el estudio. En la primera columna aparece la edad del paciente, en la segunda columna aparece la presión sistólica antes de la ingestión del medicamento y en la tercera columna aparece indicado el valor de la presión sistólica 60 minutos después de la toma del medicamento. Los 40 primeros datos corresponden a pacientes del primer grupo y los 60 restantes a pacientes del segundo grupo. También se incluye una última columna donde figura el valor del colesterol total del paciente. A partir de los datos que aparecen en la hoja de cálculo, se pide contestar a las siguientes cuestiones:

	Grupo	Edad	Presion sistólica antes	Presión sistólica después	Colesterol total
0	1	30	126	118	191
1	1	25	114	110	174
2	1	27	110	110	175
3	1	34	108	105	173
4	1	35	131	132	200
...	...	...	...	...	...
95	2	54	117	100	194
96	2	50	148	141	239
97	2	54	121	115	196
98	2	52	154	155	270
99	2	46	120	116	214

100 rows × 5 columns

Variables:

```
['Grupo', 'Edad', 'Presion sistólica antes', 'Presión sistólica después', 'Colesterol total']
```

Para esta tarea voy a utilizar el lenguaje de programación Python.

**a) Obtener, usando algún programa estadístico, las medidas de centralización y dispersión para cada uno de los dos grupos de control de la variable (grupo 1 y grupo 2) que mide la presión sistólica antes de la toma del medicamento e indica si la media en cada uno de estos grupos puede considerarse representativa a partir de los datos obtenidos.**

### GRUPO 1-JOVENES

Media: 123.42  
Mediana: 124.50  
Moda: [121, 125, 128, 131]  
Rango: 37.00  
Varianza: 81.48  
Desviación estándar: 9.03

### GRUPO 2-ADULTOS

Media: 127.48  
Mediana: 126.00  
Moda: [121]  
Rango: 58.00  
Varianza: 192.90  
Desviación estándar: 13.89

La media de ambos grupos es representativa ya que la diferencia entre las medias de ambos grupos es de 4.06. Además, la desviación estándar de ambos grupos es relativamente baja, lo que indica que los datos están agrupados alrededor de la media. Sin embargo, es importante tener en cuenta que la media no siempre es la mejor medida de tendencia central, especialmente si los datos están sesgados. En ese caso, la mediana puede ser una mejor medida de tendencia central.

**b) Estudiar la simetría y la curtosis del nivel de presión sistólica en los pacientes del segundo grupo para cada una de las mediciones de la hipertensión que aparecen en la tabla.**

Medición: Presion sistólica antes  
Coeficiente de asimetría: -0.1332  
Coeficiente de curtosis: -0.6514  
Medición: Presión sistólica después  
Coeficiente de asimetría: -0.1889  
Coeficiente de curtosis: 0.2408

Los coeficientes de asimetría y curtosis son medidas que describen la forma de una distribución. El coeficiente de asimetría mide la simetría de la distribución, mientras que el coeficiente de curtosis mide la concentración de los datos alrededor de la media. En el caso de la medición de la presión sistólica antes, el coeficiente de asimetría es -0.1332, lo que indica que la distribución es ligeramente asimétrica a la izquierda. El coeficiente de curtosis es -0.6514, lo que indica que la distribución es relativamente plana en comparación con una distribución normal. Por lo tanto, la distribución es platicúrtica.

En el caso de la medición de la presión sistólica después, el coeficiente de asimetría es -0.1889, lo que indica que la distribución es ligeramente asimétrica a la izquierda. El coeficiente de curtosis es 0.2408, lo que indica que la distribución es más puntiaguda que una distribución normal. Por lo tanto, la distribución es mesocúrtica.

**c) Indicar para cada una de las variables relacionadas con la medición de la presión sistólica que aparecen en el fichero el valor de los cuartiles y su significado y obtener el box- plot ( diagrama de cajas) correspondiente. Estudiar la presencia de valores atípicos.**

Medición: Presion sistólica antes  
min 96.0  
25% 118.0  
50% 125.0  
75% 135.0  
max 154.0

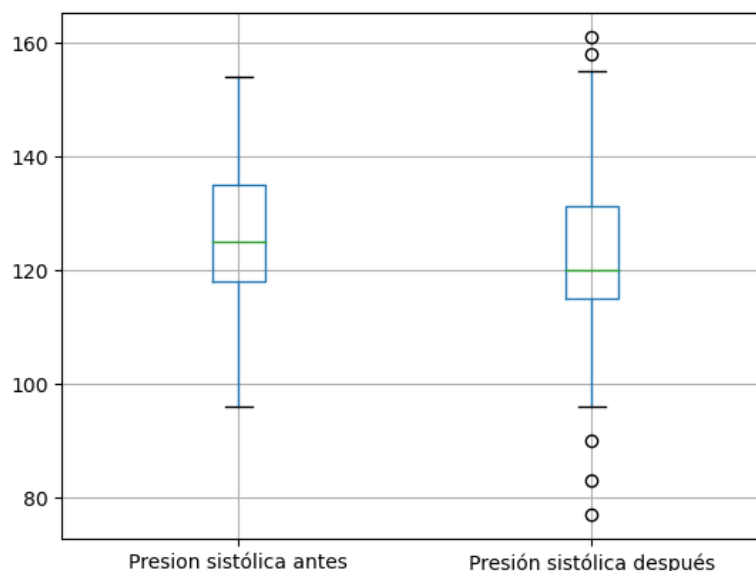
Medición: Presión sistólica después  
min 77.00  
25% 115.00  
50% 120.00  
75% 131.25  
max 161.00

Según los resultados obtenidos, se puede observar que:

La variable presión sistólica antes tiene un Q1 de 118, un Q2 de 125 y un Q3 de 154. Esto significa que el 25% de los datos son menores o iguales a 118, el 50% de los datos son menores o iguales a 125 y el 75% de los datos son menores o iguales a 154.

En el caso de la medición de la presión sistólica antes, el rango intercuartil (RIC) es 17.0. Por lo tanto, podemos identificar los valores atípicos utilizando la siguiente regla:

Valor atípico <  $Q1 - 1.5 \times RIC$  o valor atípico >  $Q3 + 1.5 \times RIC$  Valor atípico < 90.5 o valor atípico > 162.



El diagrama de cajas muestra que no hay valores atípicos para la presión sistólica antes de la toma de medicamentos. La variable presión sistólica después de la toma del medicamento, tiene un Q1 de 115, un Q2 de 120 y un Q3 de 131.25. Esto significa que el 25% de los datos son menores o iguales a 115, el 50% de los datos son menores o iguales a 120 y el 75% de los datos son menores o iguales a 131.25.

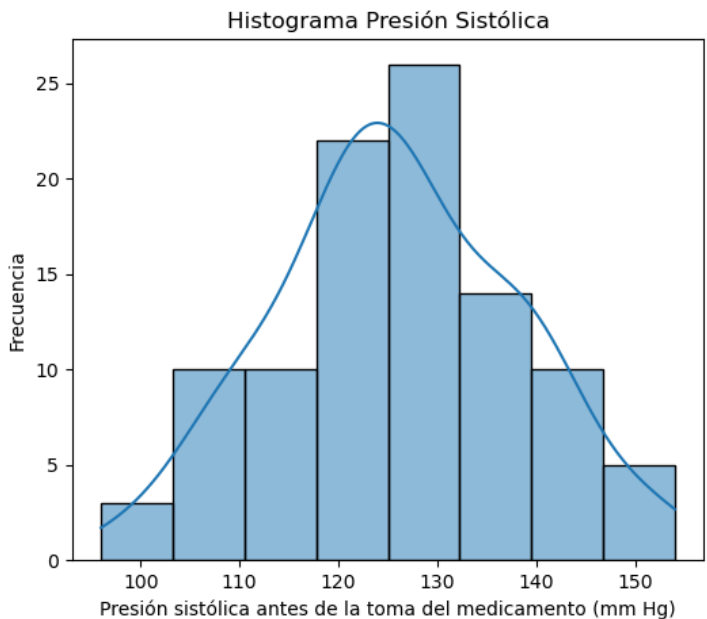
En el caso de la medición de la presión sistólica después, el rango intercuartil (RIC) es 16.25. Por lo tanto, podemos identificar los valores atípicos utilizando la siguiente regla:

Valor atípico <  $Q1 - 1.5 \times RIC$  o valor atípico >  $Q3 + 1.5 \times RIC$  Valor atípico < 89.375 o valor atípico > 157.875

Los valores atípicos de la medición de la presión sistólica después son [158, 161, 83, 90, 77]. Los valores atípicos se representan como puntos fuera del diagrama de cajas.

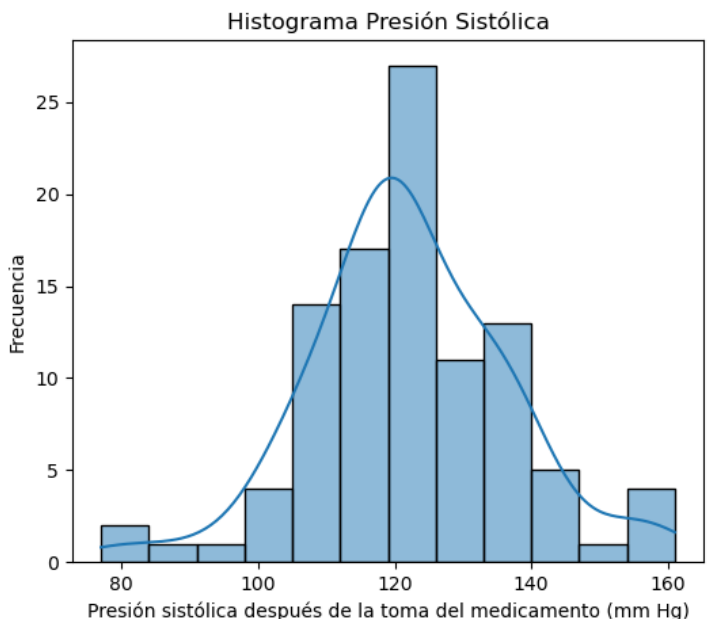
Es importante tener en cuenta que los valores atípicos pueden ser causados por errores en la recolección de datos o por la presencia de valores extremos en la población. Por lo tanto, es importante analizar cuidadosamente los valores atípicos y determinar si deben ser excluidos del análisis o si son datos importantes que deben ser conservados.

**d) Estudiar la normalidad de los datos de las variables relacionadas con la medición de la presión sistólica.**



Estadístico de prueba: 0.9919  
p-valor: 0.8157

El estadístico de prueba, de la prueba de normalidad de Shapiro-Wilk para los datos proporcionados es 0.9919 y el p-valor es 0.8157. Como el valor p es mayor que 0.05, no hay suficiente evidencia para rechazar la hipótesis nula de que los datos provienen de una distribución normal. Por lo tanto, podemos asumir que los datos siguen una distribución normal.



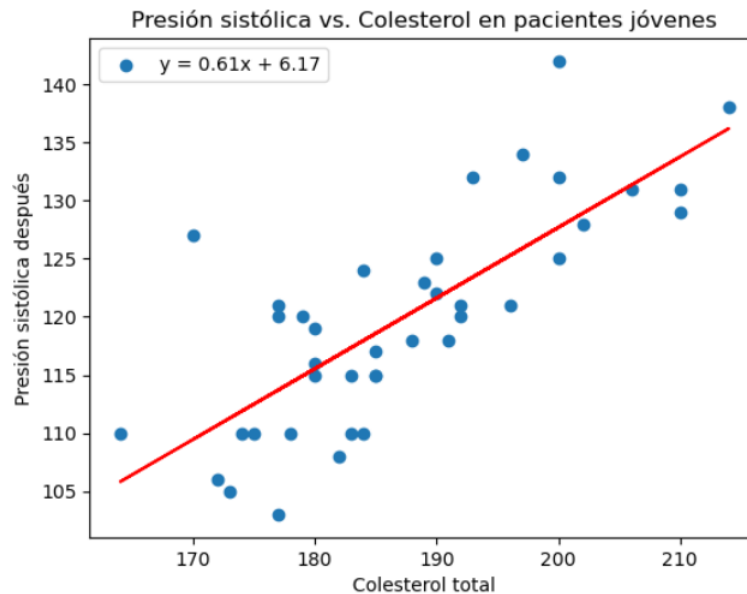
Estadístico de prueba: 0.9789  
p-valor: 0.1087

Como  $p\text{-valor} > 0.05$ , asumimos que los datos siguen una distribución normal. por lo tanto no rechazamos la hipótesis nula de normalidad. Es importante tener en cuenta que la prueba de normalidad de Shapiro-Wilk es sensible al tamaño de la muestra. Por lo tanto, es posible que la prueba no detecte desviaciones de la normalidad en muestras pequeñas, incluso si los datos no siguen una distribución normal. En ese caso, es recomendable utilizar otras pruebas de normalidad o medidas de tendencia central y dispersión para analizar los datos.

## Ejercicio 2

Con los datos del fichero anterior, se quiere estudiar la relación existente entre la presión sistólica después de la toma del medicamento y el colesterol total del paciente en los pacientes jóvenes (grupo 1).

a) Estudiar la relación lineal existente entre estas dos variables de estudio.



La ecuación de la línea de regresión es  $y = 0.61x + 6.17$ . Esto significa que por cada unidad de aumento en el colesterol total, la presión sistólica después de la toma del medicamento aumenta en promedio 0.61 unidades. El intercepto de 6.17 indica que cuando el colesterol total es cero, la presión sistólica después de la toma del medicamento es de 6.17 unidades. La pendiente de 0.61 indica que la relación entre el colesterol total y la presión sistólica después de la toma del medicamento es positiva y fuerte. Es importante tener en cuenta que la correlación no implica causalidad. Es decir, no podemos concluir que el colesterol total causa un aumento en la presión sistólica después de la toma del medicamento solo porque están correlacionados.

b) Obtener un modelo lineal que explica la presión sistólica del paciente joven a los 60 minutos de ingerir el medicamento en función de su colesterol total y realizar la estimación para un paciente del grupo 1 ( joven) cuyo colesterol total es 105 mg/dL.

El modelo lineal ajustado es:  $y = 0.61x + 6.17$

Presión sistólica estimada para un paciente joven con un colesterol total de 105 mg/dL es: 69.97

c) ¿Qué tanto por ciento de la presión sistólica del paciente joven a los 60 minutos de ingerir el medicamento no queda explicado por el anterior modelo? ¿Cómo podrías mejorar el modelo?

$R^2$ : 0.5885

Proporción no explicada: 41.15%

En este caso, el coeficiente de determinación es 0.5885, lo que indica que el modelo explica el 58.85% de la variabilidad en la presión sistólica después de la toma del medicamento en función del colesterol total del paciente. La proporción no explicada es del 41.15%, lo que indica que hay otros factores que contribuyen a la variabilidad en la presión sistólica después de la toma del medicamento que no están incluidos en el modelo. No podemos concluir que el colesterol total por sí solo causa un aumento en la presión sistólica después de la toma del medicamento, una manera de mejorar el modelo sería introduciendo nuevas variables.

d) Si aumentásemos el colesterol de un paciente en 5 mg/dL ¿Qué variación experimentaría su presión sistólica después de 60 minutos de ingerir el medicamento?

Variación de la Presión Sistólica al Incrementar el Colesterol en 5 mg/dL: 3.04

Si aumentáramos el colesterol de un paciente en 5 mg/dL, la variación de la presión sistólica después de 60 minutos de ingerir el medicamento sería de 3.04 unidades. Es importante tener en cuenta que esta es solo una estimación y que hay otros factores que pueden influir en la presión sistólica después de la toma del medicamento que no están incluidos en el modelo.

### Ejercicio 3

**a) Se quiere estudiar si se puede admitir que la presión sistólica media en el momento de la ingestión de la población adulta (grupo 2) es 130 mm de Hg. Obtener el intervalo de confianza al 95% y al 99% para el nivel medio de presión sistólica antes de la toma del medicamento en el grupo de los adultos y posteriormente contesta a la cuestión planteada con los resultados obtenidos o mediante un contraste de hipótesis.**

Para obtener el intervalo de confianza al 95% y al 99% para el nivel medio de presión sistólica antes de la toma del medicamento en el grupo de los adultos, se puede utilizar la fórmula del intervalo de confianza para la media poblacional.

Intervalo de confianza al 95%: (123.97, 131.0)  
Intervalo de confianza al 99%: (122.86, 132.1)

Como el intervalo de confianza al 95% o al 99% para el nivel medio de presión sistólica antes de la toma del medicamento en el grupo de los adultos incluye el valor de 130 mm Hg, entonces no rechazamos la hipótesis nula y podemos concluir que la presión sistólica media en el momento de la ingestión de la población adulta es 130 mm Hg.

**b) Obtener el intervalo de confianza al 95% para la diferencia de medias en la presión sistólica entre adultos y jóvenes después de la ingestión del medicamento. ¿Se puede concluir que después de la ingesta del medicamento la presión sistólica media de la población es distinta dependiendo de la edad?**

Para obtener el intervalo de confianza al 95% para la diferencia de medias en la presión sistólica entre adultos y jóvenes después de la ingestión del medicamento, se puede utilizar la fórmula del intervalo de confianza para la diferencia de medias poblacionales.

Intervalo de confianza al 95% para la diferencia de medias en la presión sistólica entre adultos y jóvenes después de la ingestión del medicamento: (-1.7168454311660994, 8.816845431166094)

El intervalo de confianza al 95% para la diferencia de medias en la presión sistólica entre adultos y jóvenes después de la ingestión del medicamento es (-1.72, 8.82). Esto significa que con un nivel de confianza del 95%, la verdadera diferencia en la presión sistólica entre adultos y jóvenes después de la ingestión del medicamento se encuentra entre -1.72 y 8.82.

Dado que el intervalo de confianza incluye el valor cero, no se puede concluir que después de la ingesta del medicamento la presión sistólica media de la población es distinta dependiendo de la edad.

**c) Se quiere estudiar la proporción de la población con una presión sistólica inicial igual o superior a 130 mm de Hg (prehipertensión). A partir de la muestra del fichero (tomando todos los datos de presión sistólica antes de la toma del medicamento) obtener un intervalo de confianza al 99% de la proporción de la población con hipertensión y contrastar la hipótesis que el porcentaje de la población con presión sistólica superior o igual a 130 mm de Hg es 0,30 con nivel de significación del 5%.**

Para obtener el intervalo de confianza al 99% para la proporción de la población con una presión sistólica inicial igual o superior a 130 mm Hg, se puede utilizar la fórmula del intervalo de confianza para una proporción poblacional.

Intervalo de confianza al 99% para la proporción de la población con una presión sistólica inicial igual o superior a 130 mm Hg: (0.2271, 0.4729)

Para contrastar la hipótesis de que el porcentaje de la población con presión sistólica superior o igual a 130 mm de Hg es 0.30 con un nivel de significación del 5%, se puede utilizar la prueba de hipótesis para una proporción poblacional. La hipótesis nula es que la proporción poblacional es igual a 0.30, y la hipótesis alternativa es que la proporción poblacional es distinta de 0.30. El valor crítico para un nivel de significación del 5% es 1.96. Si el estadístico de prueba  $z$  calculado a partir de la muestra cae fuera del intervalo (-1.96, 1.96), se rechaza la hipótesis nula.

$z$

1.0910894511799616

Como el valor calculado de  $z$  es 1.091, entonces el estadístico de prueba cae dentro del intervalo (-1.96, 1.96). Por lo tanto, no se puede rechazar la hipótesis nula y no se puede concluir que la proporción de la población con presión sistólica superior o igual a 130 mm de Hg es significativamente diferente de 0.30.

**d) (VOLUNTARIO) Por último, se quiere estudiar la eficacia del medicamento en la población adulta. ¿Existe variación significativa de la presión sistólica después de la toma del medicamento en la población del grupo 2? Plantea el**

correspondiente contraste de hipótesis considerando un nivel de significación del 5%. Ayuda: Para contestar a la pregunta has de considerar la series de datos obtenidas a partir de las diferencias entre la presión sistólica antes de la toma y la presión sistólica al cabode 60 minutos en el grupo de los adultos (contraste de muestras emparejadas).

Para contrastar la hipótesis de que no hay variación significativa de la presión sistólica después de la toma del medicamento en la población del grupo 2, se puede realizar un contraste de hipótesis para muestras emparejadas. La hipótesis nula es que la diferencia media entre la presión sistólica antes de la toma y la presión sistólica al cabo de 60 minutos después de la toma del medicamento es igual a cero, mientras que la hipótesis alternativa es que la diferencia media es distinta de cero. El nivel de significación es del 5%.

```
from scipy.stats import ttest_rel

# Calcular la diferencia entre la presión sistólica antes de la toma y la presión sistólica al cabo de 60 minutos después de la
diferencias = adultos['Presión sistólica después'] - adultos['Presion sistólica antes']

# Realizar el contraste de hipótesis
alpha = 0.05
t_statistic, p_value = ttest_rel(adultos['Presión sistólica después'], adultos['Presion sistólica antes'])
if p_value < alpha:
    print("Se rechaza la hipótesis nula")
else:
    print("No se puede rechazar la hipótesis nula")
```

Se rechaza la hipótesis nula

p\_value

0.000363117754899273

Dado que el p-valor obtenido es 0.0004, que es menor que el nivel de significación del 5%, se rechaza la hipótesis nula. Por lo tanto, se puede concluir que hay una variación significativa de la presión sistólica después de la toma del medicamento en la población adulta (grupo 2).