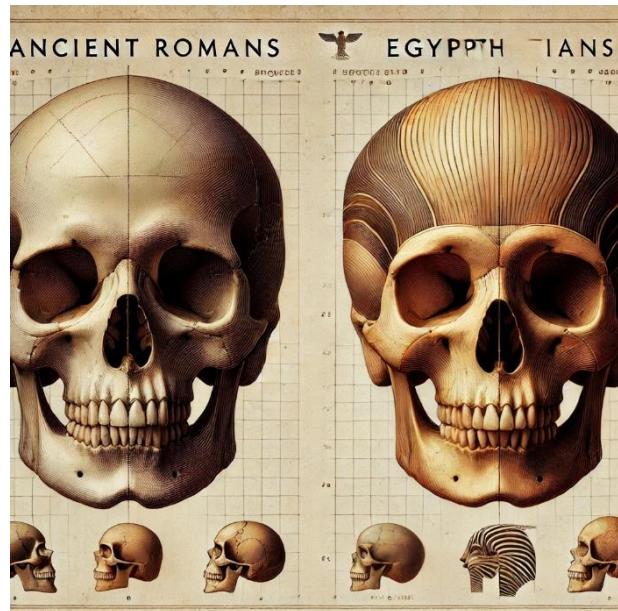


17-12-2024

Anchura de los cráneos de un yacimiento arqueológico para dos periodos temporales diferentes

ANÁLISIS DE RESULTADOS



Jacobo Álvarez Gutiérrez
MOD6 - ESTADÍSTICA - UCM

Contenido

Ejercicio 1	2
Apéndice Ejercicio 1b	5
Ejercicio 2	6
Apéndice Ejercicio 2a	8
Apéndice al Ejercicio 2b	10
Conclusiones	11

Ejercicio 1

La siguiente tabla contiene, en un editable Excel, dos variables: la primera es dicotómica con valores 1 (predinástico temprano) y 2 (predinástico tardío) y la segunda contiene la anchura de cráneos (mm) encontrados en un yacimiento arqueológico. La idea es analizar si existen diferencias en la longitud de la anchura de los cráneos egipcios a medida que pasa el tiempo. Creo que mayoritariamente tenemos una idea de que las cabezas egipcias son más alargadas y cuando ya llegamos a los romanos son más redondeadas. El cine se ha encargado de hacer muy gráfico todo esto.

Se pide:

- a) Obtener con Python las diferentes medidas de centralización y dispersión, asimetría y curtosis estudiadas. Así mismo, obtener el diagrama de caja y bigotes. Se debe hacer por separado para la submuestra de los cráneos del predinástico temprano y para la submuestra de los del predinástico tardío. Comentar los resultados obtenidos. Estos comentarios son obligatorios.

Todos los resultados que se muestran a continuación se obtuvieron empleando el Jupyter Notebook adjunto a este documento en la entrega de la tarea.

Medida	Predinástico Temprano	Predinástico Tardío
0	count	30.0
1	mean	131.533333
2	std	0.819307
3	min	130.0
4	25%	131.0
5	50%	131.5
6	75%	132.0
7	max	134.0
8	Moda	131, 132
9	Rango	4.0
10	Varianza	0.671264
11	Coeficiente Pearson	0.006229
12	Coeficiente Fisher	0.645941
13	Coeficiente Curtosis	-0.280029

FIGURA 1. Medidas de centralización y dispersión para las submuestras de cada periodo temporal.

Estos resultados numéricos ya nos aportan alguna herramienta para obtener conclusiones de ambas muestras. En primer lugar, nótese que la media del predinástico temprano resulta ligeramente menor que la del predinástico tardío. Esta tendencia también se refleja en el valor de la mediana, que parece ser visiblemente mayor en el período tardío. Por otra parte, vemos que las medidas de desviación (varianza y desviación estándar) son también ligeramente mayores en el predinástico tardío, indicando por ende una mayor dispersión de los datos en este periodo con respecto al valor medio.

Respecto a la forma de esta distribución, vemos del análisis del coeficiente de curtosis que el predinástico temprano sugiere una distribución más leptocúrtica que la distribución del predinástico tardío, esto es, con valores más próximos al valor medio y colas ligeramente más extensas que en una distribución normal.

Finalmente, en lo que respecta a su simetría, vemos que tanto los coeficientes de Fisher como de Pearson insinúan que ambas distribuciones tienen una ligera asimetría con sesgo positivo, siendo un poco más acentuada en el Predinástico Temprano. Esto quiere decir que, aunque la mayoría de los valores están concentrados por debajo del valor medio, existen algunos mayores a este que prolongan la cola de la distribución hacia valores mayores que la media.

Todas las conclusiones previas se pueden verificar a partir de las representaciones gráficas de los histogramas y los diagramas de caja y bigotes obtenidos para cada una de las muestras y que se muestran a continuación:

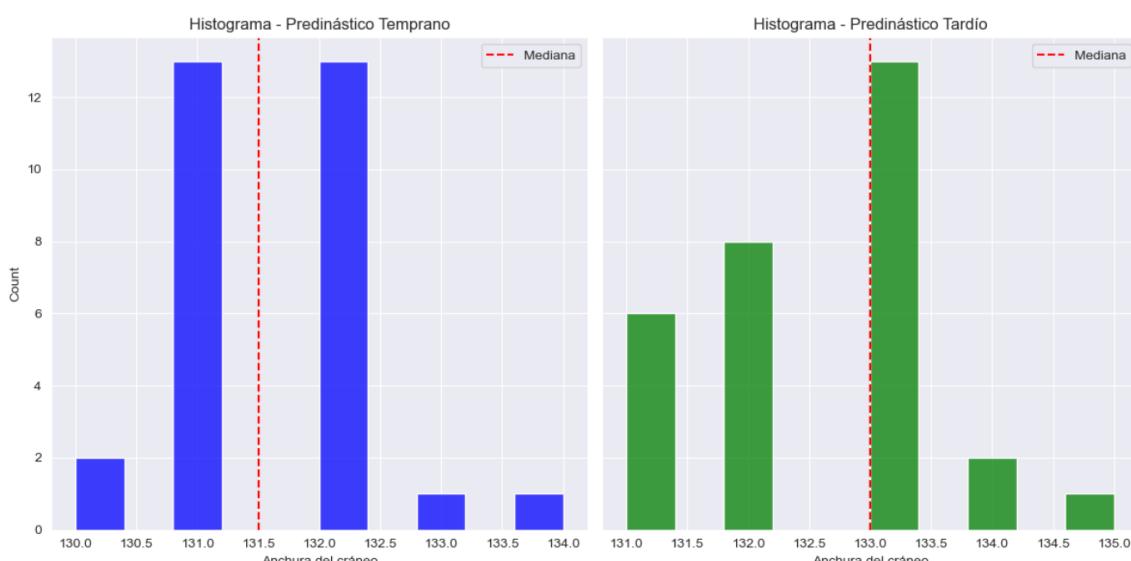


FIGURA 2. Histograma de cada una de las submuestras según su época histórica.

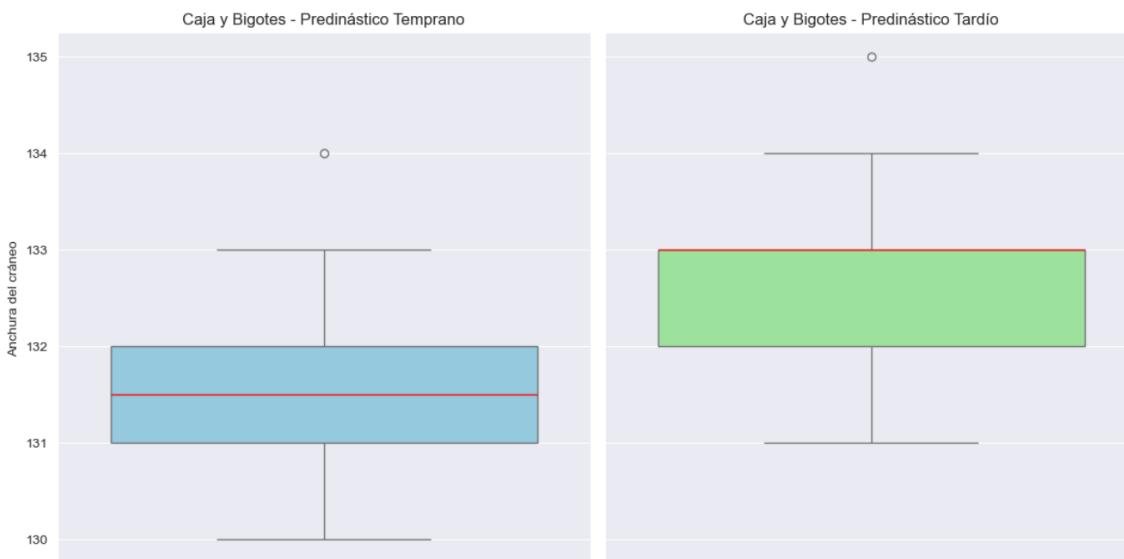


FIGURA 3. Diagrama de caja y bigotes de cada una de las submuestras según su época histórica.

En efecto, vemos en la representación gráfica de los histogramas esa tendencia de desplazamiento hacia valores mayores de anchuras craneales en el predinástico tardío, tal y como insinuaban los estadísticos centrales. De forma análoga, el diagrama de cajas muestra una mayor variabilidad en la muestra tardía, así como una mediana mucho más elevada, reflejando de nuevo esta tendencia augurada en los cálculos numéricos.

En conclusión, un primer análisis de los estadísticos principales para las muestras analizadas sugiere que existe una pequeña diferencia en las anchuras craneales entre ambos períodos temporales.

- b) Determinar si cada una de las dos submuestras sigue una distribución normal utilizando el test de Kolmogorov-Smirnov.

El test de Kolmogorov – Smirnov que se aplica a continuación se resume en el siguiente test de hipótesis:

$$\begin{cases} H_0 \rightarrow \text{La muestra } i - \text{ésima sigue una distribución normal} \\ H_1 \rightarrow \text{La muestra } i - \text{ésima no sigue una distribución normal} \end{cases}$$

Tras el cálculo del estadístico y el pvalor para cada una de las submuestras asociada a cada época histórica, se obtuvo el siguiente resultado:

```
Los resultados para la submuestra del predinástico temprano son:  
Estadístico: 0.2460, Valor p: 0.0438  
La muestra no sigue una distribución normal
```

```
Los resultados de la submuestra del predinástico tardío son:  
Estadístico: 0.2381, Valor p: 0.0557  
La muestra sigue una distribución normal
```

FIGURA 4. Resultados del test de Kolmogorov-Smirnov obtenidos con Python

Tal y como podemos apreciar de dicho análisis, parece ser que la submuestra asociada al predinástico temprano no sigue una distribución normal, mientras que la submuestra asociada al predinástico tardío sí que verifica dicha propiedad. Nótese que esta conclusión es consistente con el análisis realizado en el apartado (a), donde se apreciaba una mayor asimetría con sesgo positivo y una mayor curtosis para el primer periodo en comparación con el segundo.

Sin embargo, es de vital importancia enfatizar el hecho de que no podemos otorgar una fiabilidad absoluta a este análisis. En efecto, se trata de muestras donde la variable representada toma valores continuos y se está efectuando el test para comprobar la normalidad de la distribución, lo cual parece adecuado. Sin embargo, al tratarse de una muestra muy pequeña (30 registros en cada submuestra), la potencia del test se ve bastante limitada.

Apéndice Ejercicio 1b

Otro test más adecuado en una situación como esta sería el test de Shapiro-Wilk. Sin entrar en los aspectos teóricos del mismo, podemos afirmar que los resultados de su aplicación a las muestras que estamos trabajando son los siguientes:

```
Resultados para la submuestra del Predinástico Temprano:  
Estadístico W: 0.8378, Valor p: 0.0003  
La muestra no sigue una distribución normal.
```

```
Resultados para la submuestra del Predinástico Tardío:  
Estadístico W: 0.8832, Valor p: 0.0033  
La muestra no sigue una distribución normal.
```

FIGURA 5. Resultados del Test de Shapiro-Wilk obtenido con Python

Tal y como se puede apreciar después de este análisis, parece ser que, a diferencia de lo que sugería el test de Kolmogorov-Smirnov, ninguna de las dos muestras sigue una distribución normal. Es decir, este segundo resultado sugiere

que las desviaciones de la normalidad son, en efecto, pequeñas pero significativas. Dicho de otro modo, parece ser que, por pequeñas desviaciones, los datos de la anchura de los cráneos en estos dos períodos no se ajustan a una distribución normal en ninguno de los casos.

Nótese que esto tiene especial relevancia para la resolución del segundo ejercicio, donde deberá asumirse una normalidad en la distribución de ambas submuestras para la correcta aplicación de los análisis pertinentes. Así, asumiremos dicha condición de normalidad, pero a sabiendas de que los resultados no serán del todo fiables.

Ejercicio 2

- a) Con los mismos datos del ejercicio anterior, obtener un intervalo de confianza (de nivel 0.9, de nivel 0.95 y de nivel 0.99) para la diferencia entre las medias de la anchura de la cabeza en ambos períodos históricos. Interpretar los resultados obtenidos y discutirlos en función del test de normalidad del ejercicio anterior. La interpretación debe ser rigurosa desde el punto de vista estadístico y también marcada por el story telling, es decir, comprensible desde el punto de vista de las variables respondiendo a la pregunta ¿en qué época la cabeza era más ancha?

A continuación, se calcularán tres intervalos de confianza con diferente nivel de significación para la diferencia de medias entre ambas muestras. En primer lugar, deberíamos garantizar la independencia de las muestras para poder aplicar correctamente este análisis. Sin embargo, en base al enunciado, esta hipótesis la podemos asumir como cierta.

El siguiente paso en el análisis será tratar de determinar si existe o no alguna diferencia entre las varianzas poblacionales de cada muestra. Para ello, se realizará un intervalo de confianza al 90% para analizar la diferencia entre estas. Sin embargo, aunque veíamos en el test de Shapiro-Wilk que ninguna de las muestras se termina de asemejar a una distribución normal, vamos a asumir esta condición por simplicidad. Al finalizar las conclusiones de este apartado asumiendo normalidad, se hará un pequeño esbozo del procedimiento teórico (estadística no paramétrica) adecuado para una situación como esta.

Asumiendo entonces independencia entre las muestras y una distribución normal de cada una de ellas, podemos plantear el siguiente test de hipótesis:

$$\begin{cases} H_0 \rightarrow \text{Las muestras poseen varianzas poblacionales iguales } S_1 = S_2 \\ H_1 \rightarrow \text{Las muestras poseen varianzas poblacionales diferentes } S_1 \neq S_2 \end{cases}$$

El resultado de este se refleja en la siguiente representación gráfica:

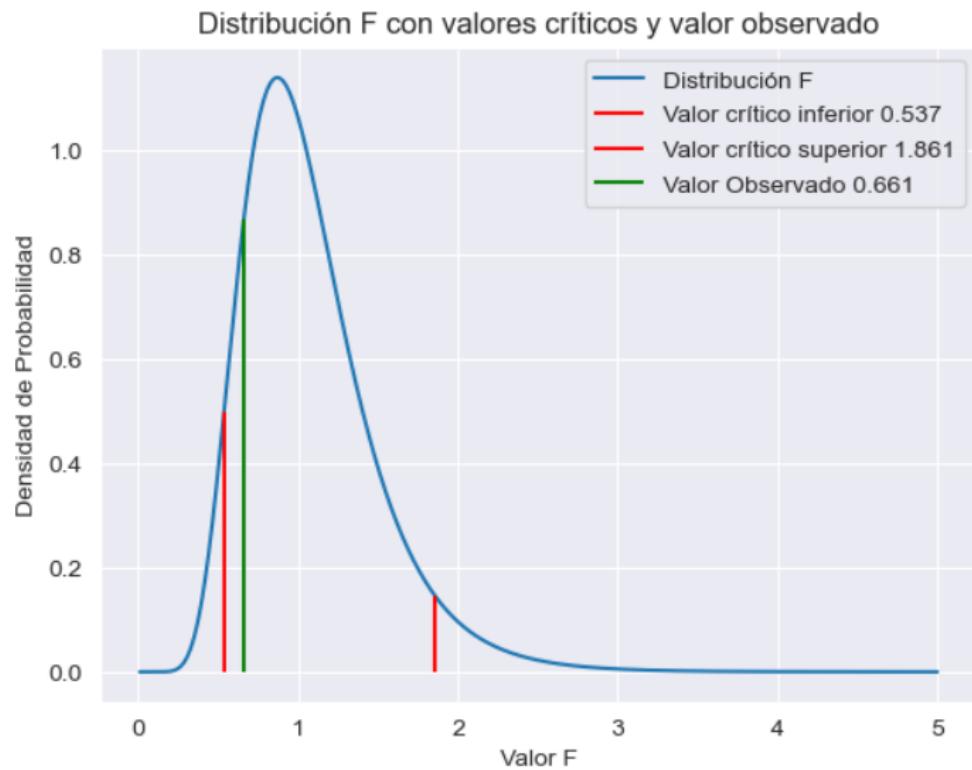


FIGURA 6. Resultado del F Test para la diferencia de varianzas obtenido con Python.

Como podemos observar, el valor calculado para el estadístico cae dentro de la región de aceptación de la hipótesis nula o, lo que es lo mismo, podemos afirmar con una confianza del 90% que no nos estamos equivocando al asumir que no existe diferencia entre las varianzas poblacionales de las muestras asociadas a cada periodo histórico.

A colación del análisis previo y el resultado inmediatamente anterior, podemos asumir independencia y normalidad en las muestras, así como que sus varianzas son desconocidas pero iguales entre sí. Esto nos da pie a realizar un estudio para la diferencia entre las medias poblacionales de cada muestra empleando intervalos de confianza para distintos niveles de significación. Los resultados obtenidos fueron los siguientes:

$$\begin{cases} IC_{90\%} = [-1.3365, -0.5301] \\ IC_{95\%} = [-1.4162, -0.4505] \\ IC_{99\%} = [-1.5758, -0.2909] \end{cases}$$

Como se puede observar, el valor 0 no pertenece a ninguno de los intervalos de confianza calculados. Por ende, los datos evidencian una discrepancia notable entre los valores de las medias poblacionales de cada muestra. Es decir, parece que la diferencia en la anchura de los cráneos para ambos períodos históricos no

es pura aleatoriedad, sino que existe una diferencia significativa, siendo mayor la de los cráneos del predinástico tardío.

Apéndice Ejercicio 2a

Si atendemos al resultado del test de Shapiro-Wilk, realmente estas distribuciones no parece que sigan la forma característica de una distribución normal, lo cual nos obliga a buscar métodos no paramétricos para hacer un análisis correcto de cada muestra. Una alternativa más adecuada a esta situación para comprobar la homogeneidad de varianzas sería por ejemplo el test de Levene que, de nuevo sin entrar en muchos aspectos teóricos, enfatiza el cálculo del estadístico en base a la mediana, lo cual hace el test más robusto frente a valores extremos. Tras una aplicación de dicho test se obtuvieron los siguientes resultados:

**Estadístico de Levene: 0.6195, Valor p: 0.4344
No hay evidencias suficientes para descartar la hipótesis nula.**

FIGURA 7. Resultado del Test de Levene para la diferencia de varianzas obtenido con Python.

Como podemos apreciar, también se deduce de este test el mismo resultado que con el F test. Es decir, parece que no hay evidencias suficientes para rechazar la idea de que ambas distribuciones tengan unas varianzas poblacionales idénticas.

Sabiendo esto, pero continuando con métodos estadísticos no paramétricos (para no asumir normalidad), una técnica razonable para el cálculo de intervalos de confianza para la diferencia de medias es el Bootstrap. No se analizará en detalle el fundamento teórico de esta técnica, pero podemos afirmar que los resultados obtenidos para las muestras de ambos períodos son:

$$\begin{aligned} IC_{90\%} &= [-1.3000, -0.5333] \\ IC_{95\%} &= [-1.3675, -0.5000] \\ IC_{99\%} &= [-1.5000, -0.3332] \end{aligned}$$

Nótese que, nuevamente, el 0 no pertenece a ninguno de los intervalos de confianza calculados. Por tanto, en consonancia con el estudio que asumía normalidad, parece que existe una diferencia significativa entre la media del ancho craneal para los individuos del predinástico temprano con respecto a los del predinástico tardío.

- b) Utilizar el test t para contrastar la hipótesis de que ambas medias son iguales. Explicar qué condiciones se deben cumplir para poder aplicar ese contraste.

Determinar si se cumplen. Admitiremos de forma natural la independencia entre ambas muestras, así que esa condición no hace falta comprobarla.

Para la aplicación del test t tenemos que verificar el cumplimiento de tres hipótesis. En primer lugar, la independencia de ambas muestras (lo cual recordamos del apartado anterior que ya era una condición que asumía el enunciado del ejercicio). Por otra parte, la homogeneidad de las varianzas, lo cual se ha probado con el F test en el apartado inmediatamente anterior (así como con el test de Levene). Y, finalmente, se debe verificar la normalidad en la distribución de ambas muestras.

Somos conscientes de que el cumplimiento de esta última hipótesis se ha puesto en entredicho en varias ocasiones durante el análisis de los datos realizado hasta ahora. Sin embargo, vamos a efectuar el T test asumiendo normalidad. Al terminar este análisis se hará un pequeño esbozo de un procedimiento más acertado (no paramétrico).

Como decíamos, si asumimos normalidad para los datos de ambas muestras se obtiene el siguiente resultado:

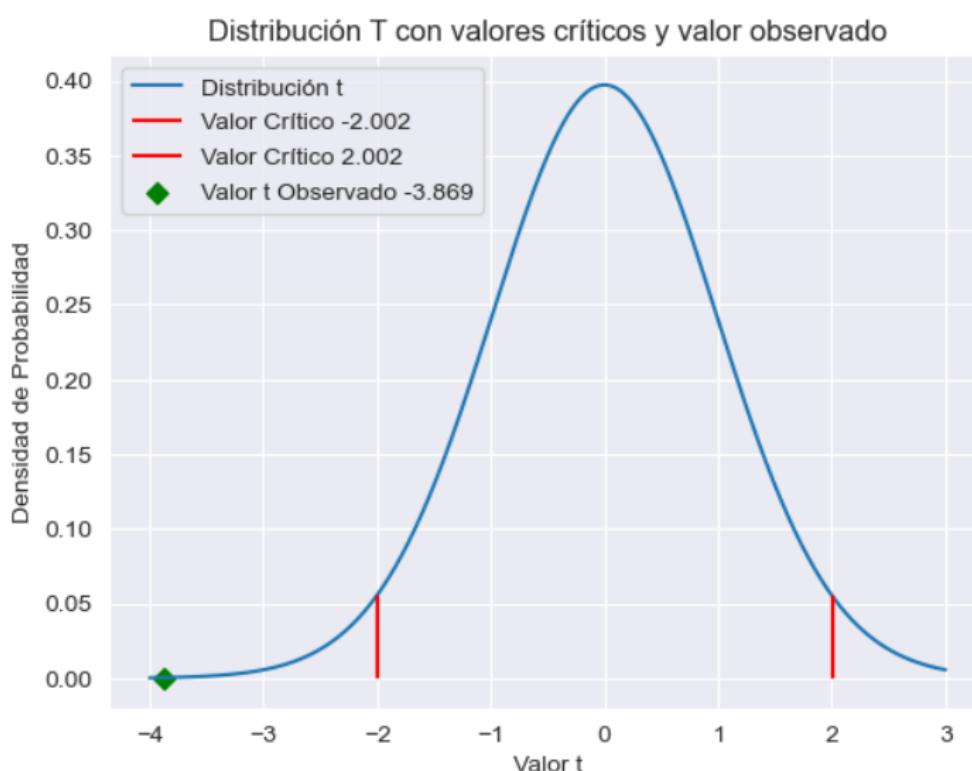


FIGURA 8. Resultado del T test para la diferencia de medias obtenido con Python.

En este caso, y en consonancia con los intervalos de confianza calculados en el apartado anterior, podemos apreciar que el valor observado cae fuera de la región de aceptación de la hipótesis nula. Es decir, no parece que haya evidencias

suficientes para aceptar una igualdad entre las medias poblacionales o, dicho de otro modo, los datos parecen volver a evidenciar que existe una diferencia significativa en el ancho del cráneo según la época histórica a la que hagan referencia (o al menos en este yacimiento arqueológico).

Apéndice al Ejercicio 2b

Si a continuación dejamos de asumir normalidad para las distribuciones de ambas muestras, es decir, si nos pasamos a métodos estadísticos no paramétricos, el análogo al T test más adecuado para esta situación sería el Test de Mann Whitney U. Sin entrar, una vez más, en aspectos técnicos de sus fundamentos teóricos, podemos decir de su implementación en Python que el resultado para la muestra usada en el ejercicio es:

```
Estadístico U: 217.5000, Valor p: 0.0003
Existen diferencias significativas entre las medianas de las muestras.
```

FIGURA 9. Resultado del Test de Mann Whitney U para la diferencia de medias obtenido con Python.

Vemos entonces, una vez más, que este test también parece evidenciar una diferencia no aleatoria en el ancho del cráneo para los individuos de las distintas épocas.

Conclusiones

El análisis estadístico realizado sobre la anchura de los cráneos encontrados en dos períodos temporales (predinástico temprano y predinástico tardío) nos ha permitido identificar diferencias significativas entre ambas muestras.

En primer lugar, los estadísticos descriptivos mostraron que la anchura media de los cráneos en el período predinástico tardío es ligeramente mayor que en el período temprano, una tendencia corroborada por la mediana y los histogramas obtenidos. La dispersión (varianza y desviación estándar) también es ligeramente mayor en el período tardío, lo cual indica una mayor variabilidad de los datos en esa muestra.

Por otra parte, el análisis de la asimetría y curtosis para ambos períodos revela que ambas distribuciones presentan una ligera asimetría con sesgo positivo, siendo un poco más leptocúrtica la del período temprano.

En lo que respecta a los test de normalidad, vimos con el test de Kolmogorov-Smirnov que solo la muestra del predinástico tardío parecía seguir una distribución normal. No obstante, como este resultado no era del todo fiable para tamaños muestrales tan pequeños, se hizo también el test de Shapiro-Wilk, el cual reveló que ninguna de las muestras parece seguir una distribución normal.

A pesar del resultado anterior, se prosiguió con métodos de estadística paramétrica asumiendo normalidad en las muestras (F test para la diferencia de varianzas y T test para la diferencia de medias, así como cálculos explícitos de intervalos de confianza), los cuales parecen reflejar una diferencia significativa en el ancho craneal de los individuos asociados a cada época.

No obstante, se ha decidido complementar el ejercicio con la implementación de algunos métodos de estadística no paramétrica (Test de Levene para la diferencia de varianzas y Test de Mann-Whitney U para la diferencia de medias, así como cálculos explícitos de intervalos de confianza Bootstrap). Sin embargo, estos métodos parecen evidenciar las mismas conclusiones que podíamos obtener con los métodos paramétricos.

En resumen, podemos concluir que existe una diferencia significativa en la anchura craneal entre los dos períodos históricos analizados, siendo mayor en el período Predinástico Tardío. Esta diferencia podría estar asociada a factores ambientales, genéticos o socioculturales, que influyeron en la evolución física de las poblaciones estudiadas.