



UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Estudios Estadísticos

Máster big data, data science & inteligencia artificial

Análisis descriptivo e inferencial de la anchura craneal en dos periodos del Egipto predinástico

Tarea de Estadística

Autor(a): Daniel Olmedo Rawlins Poveda

Profesor(a): Conrado M. Manuel García

Curso académico: 2025-2026

Madrid, 14 de enero de 2026

Índice

| | |
|---|-----------|
| 1. Introducción | 2 |
| 1.1. Contexto y motivación | 2 |
| 1.2. Objetivo y pregunta de investigación | 2 |
| 1.3. Descripción de los datos | 2 |
| 2. Metodología | 3 |
| 2.1. Plan de análisis | 3 |
| 2.2. Nivel de significación | 3 |
| 2.3. Software | 4 |
| 3. Análisis descriptivo | 4 |
| 3.1. Descripción general de las submuestras | 4 |
| 3.2. Medidas resumen por periodo | 4 |
| 3.2.1. Histogramas comparativos | 5 |
| 3.3. Análisis gráfico de la anchura craneal | 5 |
| 3.4. Evaluación descriptiva de la normalidad | 6 |
| 3.5. Interpretación descriptiva y consideraciones preliminares | 7 |
| 4. Comprobación de supuestos | 7 |
| 4.1. Normalidad | 7 |
| 4.1.1. Planteamiento de las hipótesis | 7 |
| 4.1.2. Tests de normalidad utilizados | 7 |
| 4.1.3. Resultados del contraste de normalidad | 7 |
| 4.1.4. Interpretación conjunta e implicaciones | 8 |
| 4.2. Homocedasticidad (igualdad de varianzas) | 8 |
| 4.2.1. Planteamiento de las hipótesis | 9 |
| 4.2.2. Resultados del test de Levene | 9 |
| 4.2.3. Implicaciones para el análisis inferencial | 9 |
| 5. Inferencia estadística | 9 |
| 5.1. Intervalos de confianza para $\mu_{\text{tardío}} - \mu_{\text{temprano}}$ | 9 |
| 5.1.1. Parámetro de interés | 9 |
| 5.1.2. Intervalos de confianza | 10 |
| 5.1.3. Interpretación estadística | 10 |
| 5.1.4. Interpretación contextual | 10 |
| 5.1.5. Relación con el supuesto de normalidad | 10 |
| 5.2. Contraste principal de medias | 10 |
| 5.2.1. Planteamiento de las hipótesis | 10 |
| 5.2.2. Resultados del contraste paramétrico | 11 |
| 5.3. Discusión de la validez del contraste | 11 |
| 6. Conclusiones y limitaciones | 12 |
| 6.1. Conclusión principal | 12 |
| 6.2. Limitaciones del análisis | 12 |
| Anexo A: Código y reproducibilidad | 13 |
| Anexo B: Salidas completas de los análisis | 15 |

1. Introducción

1.1 Contexto y motivación

El estudio de restos óseos procedentes de yacimientos arqueológicos constituye una fuente fundamental de información para comprender la evolución biológica y cultural de las poblaciones humanas. En particular, las características morfométricas del cráneo han sido ampliamente utilizadas para analizar posibles cambios asociados al paso del tiempo, a transformaciones culturales o a influencias ambientales.

En el caso del antiguo Egipto, existe una percepción ampliamente difundida —reforzada tanto por representaciones artísticas como por el imaginario popular— de que la morfología craneal pudo variar a lo largo de los distintos periodos históricos, tendiendo desde formas más alargadas en épocas tempranas hacia cráneos relativamente más anchos o redondeados en periodos posteriores. Este trabajo se enmarca en ese contexto, abordando el análisis cuantitativo de la anchura craneal en dos periodos del Egipto predinástico.

El uso de herramientas estadísticas permite evaluar de manera objetiva si las diferencias observadas entre periodos son atribuibles a variabilidad aleatoria o si, por el contrario, existe evidencia suficiente para sostener la presencia de un cambio sistemático en la anchura del cráneo a lo largo del tiempo.

1.2 Objetivo y pregunta de investigación

El objetivo principal de este estudio es analizar si existen diferencias estadísticamente significativas en la anchura media de los cráneos correspondientes a dos periodos históricos distintos: el predinástico temprano y el predinástico tardío.

De manera más concreta, se pretende responder a la siguiente pregunta de investigación:

¿Existen diferencias en la anchura de los cráneos entre el periodo predinástico temprano y el predinástico tardío, y en qué periodo la cabeza era, en promedio, más ancha?

Para ello, se llevará a cabo un análisis estadístico que incluye una fase descriptiva inicial, la comprobación de los supuestos necesarios para la inferencia y, finalmente, la estimación mediante intervalos de confianza y la realización de contrastes de hipótesis.

1.3 Descripción de los datos

Los datos analizados proceden de un conjunto de observaciones recogidas en un yacimiento arqueológico y se componen de dos variables:

- Una variable cualitativa dicotómica que identifica el periodo histórico al que pertenece cada cráneo, codificada como 1 para el periodo predinástico temprano y 2 para el periodo predinástico tardío.
- Una variable cuantitativa continua que representa la anchura del cráneo, medida en milímetros.

La muestra está formada por dos submuestras independientes correspondientes a cada periodo histórico, con tamaños muestrales comparables. En el conjunto de datos no se han detectado valores perdidos, lo que permite realizar el análisis sin necesidad de procedimientos adicionales de imputación o depuración.

Sobre la base de estos datos se desarrollará el análisis estadístico descrito en las secciones siguientes.

2. Metodología

2.1 Plan de análisis

El análisis estadístico se ha estructurado siguiendo un protocolo secuencial que permite, en primer lugar, explorar las características básicas de los datos y, posteriormente, aplicar técnicas inferenciales adecuadas en función de los supuestos que se cumplan. Este enfoque garantiza que las conclusiones obtenidas estén correctamente fundamentadas desde el punto de vista estadístico.

En concreto, el plan de análisis seguido ha sido el siguiente:

1. **Análisis descriptivo de los datos.** Se calcularon, por separado para cada periodo histórico, medidas de centralización y dispersión (media, mediana, varianza y desviación típica), así como medidas de forma (asimetría y curtosis). Además, se emplearon representaciones gráficas —diagramas de caja y bigotes e histogramas— con el objetivo de visualizar la distribución de la anchura craneal y detectar posibles diferencias entre periodos.
2. **Comprobación de supuestos.** Antes de aplicar técnicas inferenciales paramétricas, se evaluó la normalidad de las distribuciones correspondientes a cada submuestra mediante el test de Kolmogorov–Smirnov. Asimismo, se analizó la igualdad de varianzas entre ambos grupos mediante un test de homocedasticidad, requisito relevante para la comparación de medias.
3. **Contraste de hipótesis principal.** Con el fin de evaluar si existen diferencias estadísticamente significativas entre las medias de la anchura craneal en ambos periodos, se aplicó el test t para muestras independientes. En función del cumplimiento o no de los supuestos de normalidad y homocedasticidad, se consideró la versión adecuada del test (Student o Welch). Siguiendo las indicaciones del profesor, el contraste t se realizó incluso en el caso de que alguno de los supuestos no se cumpliera estrictamente, discutiendo posteriormente la validez de los resultados.
4. **Estimación mediante intervalos de confianza.** Se calcularon intervalos de confianza para la diferencia entre las medias de la anchura craneal en ambos periodos, utilizando niveles de confianza del 90 %, 95 % y 99 %. Estos intervalos permiten cuantificar la magnitud y la dirección de la diferencia observada, complementando la información proporcionada por el contraste de hipótesis.
5. **Interpretación crítica de los resultados.** Finalmente, los resultados obtenidos se interpretaron de manera conjunta, integrando la información procedente del análisis descriptivo, los tests de supuestos, los contrastes de hipótesis y los intervalos de confianza. Asimismo, se discutieron las principales limitaciones del análisis y las posibles implicaciones del incumplimiento de algunos supuestos estadísticos.

2.2 Nivel de significación

En todos los contrastes de hipótesis realizados se ha adoptado un nivel de significación $\alpha = 0,05$. Este valor representa la probabilidad máxima de cometer un error de tipo I que se considera aceptable en el contexto de este estudio. En el caso de los intervalos de confianza, además del nivel habitual del 95 %, se han considerado niveles del 90 % y del 99 %, tal como se solicita en el enunciado del ejercicio.

2.3 Software

El análisis estadístico se llevó a cabo utilizando el lenguaje de programación **Python**, debido a su amplia disponibilidad de librerías especializadas para el tratamiento de datos y la inferencia estadística. En particular, se emplearon las siguientes librerías:

- **NumPy** y **pandas** para la manipulación y el análisis de los datos.
- **SciPy** y **statsmodels** para la realización de pruebas estadísticas e intervalos de confianza.
- **matplotlib** y **seaborn** para la generación de las representaciones gráficas.

El código desarrollado permite reproducir íntegramente los resultados presentados en este informe y se entrega como archivo independiente, tal y como se indica en los requisitos del ejercicio.

3. Análisis descriptivo

3.1 Descripción general de las submuestras

El análisis descriptivo se realizó sobre la variable *anchura del cráneo* (en milímetros), diferenciando entre dos submuestras independientes correspondientes al periodo predinástico temprano y al periodo predinástico tardío. Cada submuestra está compuesta por 30 observaciones, lo que permite una comparación equilibrada entre ambos periodos históricos.

El objetivo de este apartado es caracterizar las principales propiedades estadísticas de ambas distribuciones antes de proceder al análisis inferencial, identificando posibles diferencias en centralización, dispersión y forma.

3.2 Medidas resumen por periodo

En la Tabla 1 se recogen las principales medidas descriptivas de la anchura craneal para cada periodo histórico.

Cuadro 1: Estadísticos descriptivos de la anchura craneal por periodo

| Periodo | <i>n</i> | Media | Mediana | Desv. típica | Asimetría | Curtosis |
|-----------------------|----------|--------|---------|--------------|-----------|----------|
| Predinástico temprano | 30 | 131.53 | 131.50 | 0.82 | 0.66 | 1.30 |
| Predinástico tardío | 30 | 132.47 | 133.00 | 1.01 | 0.20 | -0.19 |

Desde el punto de vista de la centralización, tanto la media como la mediana de la anchura craneal son mayores en el periodo predinástico tardío. En concreto, la diferencia de medias entre ambos periodos es de aproximadamente 0.93 mm, mientras que la diferencia de medianas alcanza los 1.5 mm. Estos resultados sugieren, de forma preliminar, un desplazamiento del centro de la distribución hacia valores mayores en el periodo tardío.

En cuanto a la dispersión, el periodo predinástico tardío presenta una desviación típica y una varianza superiores a las del periodo temprano, lo que indica una mayor variabilidad en las anchuras craneales durante dicho periodo. Aunque el rango total de valores es similar en ambos casos, la mayor dispersión observada en el periodo tardío será relevante al evaluar posteriormente el supuesto de homocedasticidad.

Respecto a la forma de las distribuciones, el periodo predinástico temprano muestra una asimetría positiva moderada y una curtosis claramente positiva, lo que sugiere una distribución más apuntada que la normal y con una cola más larga hacia valores altos. Por el contrario, el periodo predinástico tardío presenta una asimetría cercana a cero y una curtosis ligeramente negativa, indicando una distribución más simétrica y algo más plana que la normal.

3.2.1. Histogramas comparativos

Con el objetivo de visualizar la distribución de la anchura craneal en cada periodo y detectar posibles diferencias en la forma de las distribuciones, se elaboraron histogramas comparativos para las submuestras correspondientes al periodo predinástico temprano y al periodo predinástico tardío.

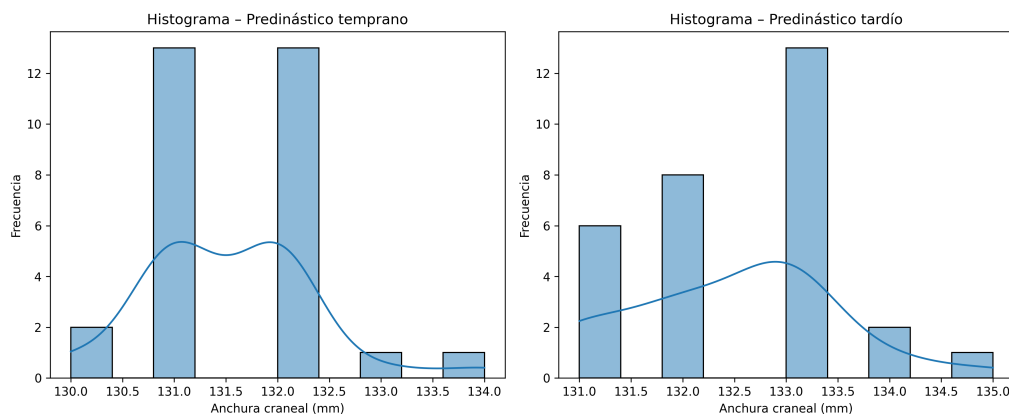


Figura 1: Histogramas comparativos de la anchura craneal por periodo

En el periodo **predinástico temprano**, la distribución de la anchura craneal presenta un patrón aproximadamente unimodal, con mayor concentración de observaciones entre 131 y 132 mm. Sin embargo, se aprecia una ligera **asimetría positiva**, coherente con el valor de asimetría estimado (0.66), lo que sugiere la presencia de una cola hacia valores altos.

En el periodo **predinástico tardío**, el histograma muestra un **desplazamiento hacia valores mayores** de anchura, con una concentración notable en torno a 133 mm. La distribución aparece más regular y próxima a la simetría, en línea con la asimetría observada (0.20). No obstante, al igual que en el periodo temprano, se observan valores en el extremo superior que podrían influir en la forma global de la distribución.

En conjunto, los histogramas sugieren un **incremento descriptivo de la anchura craneal** en el periodo tardío respecto al temprano. Aun así, dado el solapamiento parcial entre distribuciones y las diferencias en forma y dispersión, será necesario confirmar estas observaciones mediante inferencia estadística tras evaluar formalmente los supuestos de normalidad y homocedasticidad.

3.3 Análisis gráfico de la anchura craneal

Con el objetivo de complementar el análisis numérico, se realizó un análisis gráfico de la anchura craneal en ambos periodos históricos mediante diagramas de caja y bigotes. Este tipo de representación permite comparar de forma visual la centralización, la dispersión y la presencia de valores atípicos entre las submuestras.

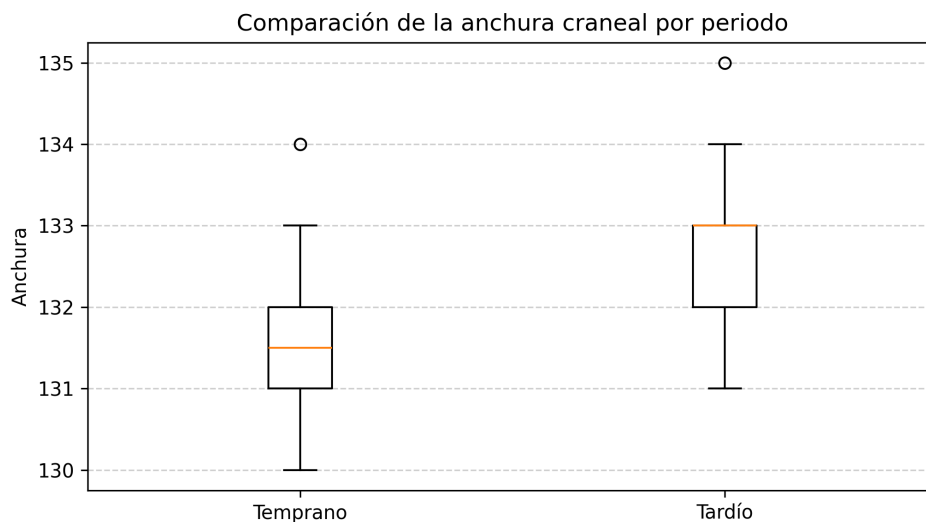


Figura 2: Diagramas de caja comparativos de la anchura craneal por periodo

El análisis visual mediante diagramas de caja confirma las diferencias observadas en las medidas de centralización. La mediana correspondiente al periodo predinástico tardío se sitúa claramente por encima de la del periodo temprano, lo que refuerza la evidencia descriptiva de un incremento de la anchura craneal a lo largo del tiempo.

Asimismo, el rango intercuartílico del periodo tardío es ligeramente mayor, lo que indica una mayor variabilidad del 50 % central de las observaciones. Los bigotes del periodo tardío se extienden hacia valores más altos de anchura, y aunque en ambos periodos se identifican valores atípicos, estos no parecen dominar la estructura general de las distribuciones.

A pesar del desplazamiento observado, existe un solapamiento parcial entre las distribuciones de ambos periodos. Este solapamiento sugiere que, aunque las diferencias descriptivas son claras, no existe una separación completa entre las anchuras craneales correspondientes a los dos periodos históricos.

3.4 Evaluación descriptiva de la normalidad

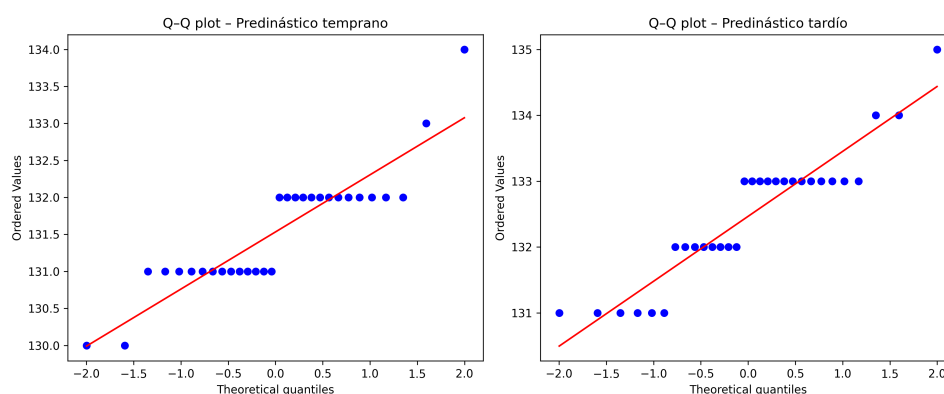


Figura 3: Diagramas de Q-Q para ver el ajuste de las submuestras

La evaluación visual de la normalidad mediante diagramas Q-Q muestra un comportamiento distinto en ambas submuestras. En el periodo predinástico temprano se observan desviaciones notables respecto a la recta teórica en los cuantiles extremos, especialmente en la cola derecha, lo que es coherente con la asimetría positiva y la curtosis elevada observadas en el análisis numérico.

Por su parte, el periodo predinástico tardío presenta un mejor ajuste global a la normalidad, con una alineación más consistente de los cuantiles empíricos a lo largo de la mayor parte del rango. No obstante, también se aprecian desviaciones en los extremos, lo que aconseja cautela a la hora de asumir una distribución normal estricta.

3.5 Interpretación descriptiva y consideraciones preliminares

En conjunto, el análisis descriptivo pone de manifiesto diferencias sistemáticas en la anchura craneal entre los periodos predinástico temprano y predinástico tardío. Los valores centrales son mayores en el periodo tardío y la variabilidad es también más elevada, lo que sugiere una posible evolución temporal hacia cráneos de mayor anchura y mayor heterogeneidad.

Sin embargo, las diferencias observadas en la forma de las distribuciones, junto con la presencia de asimetría y desviaciones respecto a la normalidad, indican que los supuestos necesarios para la aplicación de contrastes paramétricos deben ser evaluados con cuidado. Estas observaciones justifican la comprobación formal de la normalidad y la homocedasticidad en el apartado siguiente, así como la consideración de métodos alternativos en caso de que dichos supuestos no se cumplan estrictamente.

4. Comprobación de supuestos

4.1 Normalidad

Antes de aplicar contrastes inferenciales paramétricos, se evaluó el supuesto de normalidad de la variable *anchura del cráneo* en cada una de las submuestras correspondientes a los periodos predinástico temprano y predinástico tardío. La comprobación de este supuesto resulta especialmente relevante para contextualizar la validez teórica del contraste de igualdad de medias mediante el test t .

4.1.1. Planteamiento de las hipótesis

Para cada submuestra se contrastaron las siguientes hipótesis, comunes a todos los tests de normalidad aplicados:

H_0 : la muestra procede de una distribución normal, H_1 : la muestra no procede de una distribución normal.

Se adoptó un nivel de significación $\alpha = 0,05$, y los contrastes se realizaron de forma bilateral, al pretender detectar cualquier tipo de desviación respecto a la normalidad.

4.1.2. Tests de normalidad utilizados

De acuerdo con el enunciado del ejercicio y las directrices indicadas en clase, se aplicó el test de **Kolmogorov–Smirnov (KS)** con estandarización previa de los datos. Con el fin de reforzar el análisis, y dado el tamaño muestral moderado de las submuestras ($n = 30$), se complementó este contraste con los tests de **Shapiro–Wilk** y **Lilliefors**, los cuales presentan mayor potencia cuando los parámetros de la distribución normal se estiman a partir de la muestra.

Los contrastes se realizaron de manera independiente para cada periodo histórico.

4.1.3. Resultados del contraste de normalidad

En la Tabla 2 se resumen los resultados obtenidos para cada submuestra y cada uno de los tests aplicados.

Cuadro 2: Pruebas de normalidad por periodo histórico

| Periodo | Prueba | Estadístico | <i>p</i> -valor |
|-----------------------|--------------------|-------------|-----------------|
| Predinástico temprano | Kolmogorov–Smirnov | 0.2425 | 0.0489 |
| | Shapiro–Wilk | 0.8378 | 0.0003 |
| | Lilliefors | 0.2425 | 0.0010 |
| Predinástico tardío | Kolmogorov–Smirnov | 0.2350 | 0.0611 |
| | Shapiro–Wilk | 0.8832 | 0.0033 |
| | Lilliefors | 0.2350 | 0.0010 |

Para la submuestra correspondiente al **periodo predinástico temprano**, los tres contrastes aplicados arrojan *p*-valores inferiores al nivel de significación establecido. En consecuencia, se **rechaza la hipótesis nula de normalidad**, indicando que la distribución de la anchura craneal en este periodo presenta desviaciones estadísticamente significativas respecto a una distribución normal. Este resultado es coherente con lo observado en el análisis descriptivo previo, donde se detectaron asimetría positiva moderada y curtosis elevada.

En el caso del **periodo predinástico tardío**, el test de Kolmogorov–Smirnov no rechaza la hipótesis de normalidad ($p = 0,0611 > 0,05$). Sin embargo, tanto el test de Shapiro–Wilk como la corrección de Lilliefors sí rechazan la hipótesis nula ($p < 0,05$). Esta discrepancia sugiere que el test KS podría presentar menor sensibilidad frente a desviaciones moderadas de la normalidad en esta submuestra.

4.1.4. Interpretación conjunta e implicaciones

La evidencia conjunta de los contrastes de normalidad indica que el supuesto de normalidad **no se cumple de forma robusta en ninguna de las dos submuestras**. En línea con las directrices del curso, y dada la mayor potencia de Shapiro–Wilk y Lilliefors para tamaños muestrales moderados, se otorga mayor peso a estos resultados frente al test KS cuando existen discrepancias.

Este hecho tiene implicaciones directas para el análisis inferencial posterior. En particular, aunque se reportará el contraste paramétrico solicitado (test *t* para la diferencia de medias), sus resultados deberán interpretarse con cautela. Por este motivo, y con el fin de evaluar la robustez de las conclusiones, el análisis se complementará con un contraste no paramétrico (test de Mann–Whitney), tal como se desarrolla en secciones posteriores.

4.2 Homocedasticidad (igualdad de varianzas)

Además del supuesto de normalidad, la aplicación de contrastes paramétricos para la comparación de medias requiere evaluar la igualdad de varianzas entre las poblaciones analizadas, condición conocida como *homocedasticidad*. El cumplimiento de este supuesto resulta relevante para determinar la versión más adecuada del test *t* para muestras independientes.

En el presente estudio, el análisis descriptivo previo mostró diferencias en la dispersión de la anchura craneal entre ambos periodos históricos, siendo la varianza y la desviación típica mayores en el periodo predinástico tardío. Asimismo, los tests de normalidad indicaron desviaciones respecto a la normalidad, especialmente en el periodo predinástico temprano. En este contexto, resulta aconsejable emplear un contraste de igualdad de varianzas que sea **robusto frente a desviaciones de la normalidad**.

Por este motivo, se seleccionó el **test de Levene** para evaluar la homocedasticidad. A diferencia del test F clásico, el test de Levene no asume normalidad estricta y es menos sensible a la presencia de asimetría o valores extremos, lo que lo hace especialmente adecuado en situaciones como la presente.

4.2.1. Planteamiento de las hipótesis

Las hipótesis contrastadas mediante el test de Levene fueron las siguientes:

$$H_0 : \sigma_{\text{temprano}}^2 = \sigma_{\text{tardío}}^2, \quad H_1 : \sigma_{\text{temprano}}^2 \neq \sigma_{\text{tardío}}^2.$$

Se adoptó un nivel de significación $\alpha = 0,05$.

4.2.2. Resultados del test de Levene

En la Tabla 3 se presentan los resultados obtenidos al aplicar el test de Levene para evaluar la igualdad de varianzas entre ambos periodos históricos.

Cuadro 3: Test de homocedasticidad (Levene)

| Prueba | Estadístico | p -valor |
|--------|-------------|------------|
| Levene | 0.6195 | 0.4344 |

Dado que el p -valor obtenido es superior al nivel de significación establecido ($p = 0,4344 > 0,05$), **no se rechaza la hipótesis nula de igualdad de varianzas**. Por tanto, desde un punto de vista estadístico, puede considerarse plausible el supuesto de homocedasticidad entre las submuestras correspondientes a los periodos predinástico temprano y predinástico tardío.

4.2.3. Implicaciones para el análisis inferencial

El resultado del test de Levene sugiere que las varianzas de ambas submuestras no difieren de forma estadísticamente significativa. En consecuencia, desde el punto de vista de la homocedasticidad, sería admisible la aplicación del test t clásico para la comparación de medias.

No obstante, teniendo en cuenta que el supuesto de normalidad no se cumple de manera estricta en ninguna de las dos submuestras, y siguiendo un enfoque conservador y robusto, se optó por emplear el **test t de Welch** en el análisis inferencial posterior. Esta elección permite relajar el supuesto de igualdad de varianzas y proporciona resultados más fiables cuando existen desviaciones moderadas de los supuestos clásicos.

5. Inferencia estadística

5.1 Intervalos de confianza para $\mu_{\text{tardío}} - \mu_{\text{temprano}}$

Además de los contrastes de hipótesis, resulta fundamental estimar la magnitud de la diferencia entre ambos periodos históricos. Con este objetivo, se construyeron intervalos de confianza para la diferencia entre las medias poblacionales de la anchura craneal correspondientes al periodo predinástico tardío y al periodo predinástico temprano.

5.1.1. Parámetro de interés

El parámetro de interés se define como:

$$\Delta = \mu_{\text{tardío}} - \mu_{\text{temprano}},$$

donde un valor positivo de Δ indica que la anchura media del cráneo es mayor en el periodo predinástico tardío.

La estimación puntual obtenida a partir de las muestras es:

$$\hat{\Delta} = 0,933 \text{ mm.}$$

5.1.2. Intervalos de confianza

En la Tabla 4 se presentan los intervalos de confianza contruidos para tres niveles de confianza distintos, tal como se solicita en el enunciado del ejercicio.

Cuadro 4: Intervalos de confianza para la diferencia de medias ($\mu_{\text{tardío}} - \mu_{\text{temprano}}$)

| Nivel de confianza | Límite inferior (mm) | Límite superior (mm) |
|--------------------|----------------------|----------------------|
| 0.90 | 0.537 | 1.330 |
| 0.95 | 0.458 | 1.408 |
| 0.99 | 0.301 | 1.566 |

Como es esperable desde el punto de vista teórico, al aumentar el nivel de confianza los intervalos se ensanchan, reflejando una mayor incertidumbre en la estimación del parámetro poblacional.

5.1.3. Interpretación estadística

En los tres niveles de confianza considerados (90 %, 95 % y 99 %), los intervalos de confianza obtenidos no contienen el valor 0. Este hecho indica que existen diferencias estadísticamente significativas entre las medias de la anchura craneal en ambos periodos históricos.

Además, todos los intervalos son estrictamente positivos, lo que proporciona evidencia consistente de que la anchura media del cráneo es mayor en el periodo predinástico tardío que en el periodo predinástico temprano.

5.1.4. Interpretación contextual

Desde un punto de vista histórico y morfológico, los resultados sugieren que, con el paso del tiempo, los cráneos egipcios tendieron a presentar una mayor anchura. En términos cuantitativos, la anchura media del cráneo en el periodo predinástico tardío supera a la del periodo temprano en aproximadamente 0.9 mm.

El hecho de que esta diferencia se mantenga estadísticamente significativa incluso al 99

5.1.5. Relación con el supuesto de normalidad

Los intervalos de confianza se han construido mediante un enfoque paramétrico. Si bien los tests de normalidad indicaron que el supuesto de normalidad no se cumple de forma estricta, especialmente en el periodo predinástico temprano, el tamaño muestral moderado y la relativa robustez del estimador de la media permiten considerar razonable esta aproximación.

No obstante, los resultados deben interpretarse con cautela y se complementan, en secciones posteriores, con contrastes no paramétricos que permiten evaluar la robustez de las conclusiones frente al incumplimiento de los supuestos clásicos.

5.2 Contraste principal de medias

Con el fin de evaluar formalmente si existen diferencias estadísticamente significativas entre las medias de la anchura craneal correspondientes a los periodos predinástico temprano y predinástico tardío, se realizó un contraste de hipótesis para muestras independientes.

5.2.1. Planteamiento de las hipótesis

Las hipótesis contrastadas fueron las siguientes:

$$H_0 : \mu_{\text{temprano}} = \mu_{\text{tardío}}, \quad H_1 : \mu_{\text{temprano}} \neq \mu_{\text{tardío}}.$$

Se adoptó un nivel de significación $\alpha = 0,05$ y se consideró un contraste bilateral, al no presuponer a priori la dirección de la posible diferencia.

5.2.2. Resultados del contraste paramétrico

Teniendo en cuenta que el supuesto de normalidad no se cumple de forma estricta en ninguna de las dos submuestras, y siguiendo un enfoque conservador y robusto, se utilizó el **test *t* de Welch** para la comparación de medias. Este contraste no requiere igualdad de varianzas y es más robusto frente a desviaciones moderadas de los supuestos clásicos.

En la Tabla 5 se resumen los resultados obtenidos mediante el contraste paramétrico y su alternativa no paramétrica.

Cuadro 5: Contrastes de hipótesis para la comparación entre periodos

| Test | Estadístico | <i>p</i> -valor | Decisión ($\alpha = 0,05$) |
|----------------|-------------|-----------------|------------------------------|
| t de Welch | $t = 3,935$ | 0.000233 | Se rechaza H_0 |
| Mann–Whitney U | $U = 217,5$ | 0.00033 | Se rechaza H_0 |

El test *t* de Welch arroja un *p*-valor muy inferior al nivel de significación establecido, lo que conduce al rechazo de la hipótesis nula de igualdad de medias. En consecuencia, existe evidencia estadísticamente significativa de que la anchura media del cráneo difiere entre ambos periodos históricos.

5.3 Discusión de la validez del contraste

Aunque el test *t* es relativamente robusto frente a desviaciones moderadas de la normalidad, especialmente cuando los tamaños muestrales son similares, los resultados de los tests de normalidad indicaron que dicho supuesto no se cumple estrictamente, en particular en la submuestra correspondiente al periodo predinástico temprano.

El incumplimiento del supuesto de normalidad podría, en principio, afectar a la validez teórica del contraste paramétrico, influyendo en la distribución del estadístico de contraste y, por tanto, en la precisión del *p*-valor obtenido. No obstante, el tamaño muestral moderado ($n = 30$ en cada grupo) y la utilización del test de Welch mitigan parcialmente este efecto, haciendo razonable la aplicación del contraste paramétrico desde un punto de vista práctico.

Con el objetivo de evaluar la robustez de la conclusión obtenida, se aplicó de forma complementaria el **test no paramétrico de Mann–Whitney U**. Este contraste no asume normalidad y compara la posición central de las distribuciones mediante rangos, proporcionando una alternativa adecuada cuando los supuestos paramétricos no se cumplen.

El hecho de que ambos contrastes —paramétrico y no paramétrico— conduzcan al rechazo de la hipótesis nula refuerza la solidez de la conclusión inferencial. Esta concordancia sugiere que la diferencia observada en la anchura craneal entre los periodos predinástico temprano y tardío no depende del método estadístico empleado, sino que refleja una diferencia estructural entre ambas poblaciones.

En consecuencia, puede afirmarse que existe evidencia estadísticamente significativa y metodológicamente robusta de diferencias en la anchura de los cráneos entre ambos periodos históricos, siendo mayor la anchura media en el periodo predinástico tardío.

6. Conclusiones y limitaciones

6.1 Conclusión principal

El objetivo principal de este estudio fue analizar si existían diferencias estadísticamente significativas en la anchura de los cráneos egipcios entre los periodos predinástico temprano y predinástico tardío, así como determinar en qué periodo histórico la cabeza presentaba mayor anchura media.

El análisis realizado proporciona evidencia estadística sólida de que la anchura craneal difiere entre ambos periodos. Tanto el análisis descriptivo como los intervalos de confianza y los contrastes de hipótesis indican que la anchura media del cráneo es mayor en el periodo predinástico tardío. En particular, la diferencia de medias estimada es positiva y los intervalos de confianza al 90 %, 95 % y 99 % no contienen el valor cero, lo que refuerza la existencia de una diferencia real entre periodos.

Asimismo, el contraste paramétrico mediante el test t de Welch y la prueba no paramétrica de Mann–Whitney U conducen a la misma decisión inferencial: el rechazo de la hipótesis nula de igualdad entre ambos grupos. La coherencia entre ambos enfoques sugiere que la conclusión alcanzada es robusta frente al incumplimiento estricto de algunos supuestos estadísticos. En consecuencia, y en respuesta a la pregunta de investigación planteada, puede afirmarse con un alto grado de confianza que la anchura craneal es mayor en el periodo predinástico tardío que en el periodo predinástico temprano.

6.2 Limitaciones del análisis

A pesar de la solidez de los resultados obtenidos, es necesario interpretar las conclusiones a la luz de diversas limitaciones metodológicas.

En primer lugar, el supuesto de normalidad no se cumple de forma estricta en la submuestra correspondiente al periodo predinástico temprano, tal como indican los tests de Kolmogorov–Smirnov, Shapiro–Wilk y Lilliefors. Aunque se han empleado contrastes robustos, como el test t de Welch, esta desviación de la normalidad puede afectar a la precisión teórica de las inferencias paramétricas y aconseja cautela en la interpretación de los resultados.

En segundo lugar, el tamaño muestral, si bien es moderado y equilibrado entre ambos grupos ($n = 30$), limita la potencia estadística de algunos contrastes y la capacidad de detectar diferencias más sutiles en la forma de las distribuciones. Asimismo, la presencia de valores extremos, aunque no dominante, podría influir en determinadas medidas resumen.

Por otra parte, el análisis asume la independencia entre las observaciones, hipótesis razonable en el contexto del estudio pero que no puede verificarse empíricamente con la información disponible sobre el proceso de muestreo arqueológico. Del mismo modo, los resultados obtenidos se refieren exclusivamente al conjunto de cráneos analizado y a su contexto específico, lo que restringe la generalización de las conclusiones a otros yacimientos o periodos históricos.

Finalmente, aunque la coincidencia entre los enfoques paramétrico y no paramétrico refuerza la robustez de la inferencia, debe tenerse en cuenta que estos contrastes evalúan aspectos ligeramente distintos (medias frente a distribuciones o rangos), lo que introduce matices en la interpretación del efecto observado.

En conjunto, estas limitaciones no invalidan los resultados alcanzados, pero sí subrayan la necesidad de interpretarlos de forma prudente y dentro del marco metodológico en el que han sido obtenidos. No obstante, la coherencia entre descriptivos, intervalos de confianza y contrastes de hipótesis permite concluir que el análisis proporciona evidencia consistente de una mayor anchura craneal en el periodo predinástico tardío.

Anexo A: Código y reproducibilidad

A.0. Repositorio del proyecto

El código fuente completo del análisis, junto con el conjunto de datos, las figuras generadas y el informe final, se encuentra disponible en un repositorio público de GitHub.

El repositorio puede clonarse o descargarse desde la siguiente dirección:

<https://github.com/danielrawlins21/estadistico-predino-temprano-tardio>

Para clonar el repositorio localmente se puede utilizar el comando:

```
git clone https://github.com/danielrawlins21/estadistico-predino-temprano-tardio.git
```

Alternativamente, el proyecto puede descargarse como archivo comprimido (.zip) directamente desde la interfaz web de GitHub.

A.1. Entrega y contenido

Para garantizar la reproducibilidad del análisis, se entrega el código utilizado en Python junto con los ficheros necesarios para ejecutar el flujo completo (carga de datos, análisis descriptivo, validación de supuestos, inferencia estadística y generación de figuras).

La entrega incluye:

- Un script o notebook principal con el análisis completo (.py o .ipynb).
- Carpeta con salidas gráficas (figuras) utilizadas en el informe.
- Archivo `requirements.txt` con el entorno de dependencias (librerías) para replicar los resultados.

A.2. Estructura recomendada del proyecto

La estructura de carpetas utilizada (o recomendada) para el proyecto es la siguiente:

```
craneos-analisis/
|-- data/
|   |-- data.xlsx
|
|-- scripts/
|   |-- analisis_craneos.py
|
|-- outputs/
|   |-- figures/
|       |-- histogramas_submuestras.png
|       |-- boxplot_temprano.png
|       |-- boxplot_tardio.png
|       |-- boxplot_comparativo.png
|       |-- qqplot_comparativo.png
|
|-- docs/
|   |-- informe.tex
|
|-- requirements.txt
`-- README.md
```

Esta organización permite separar claramente los datos, el código, las salidas generadas y el documento final.

A.3. Dependencias y entorno

El análisis fue realizado en `Python` utilizando librerías estándar de análisis estadístico y visualización. Las dependencias exactas del entorno se encuentran especificadas en el fichero `requirements.txt` (entregado junto al proyecto).

A.4. Instrucciones de ejecución

Para reproducir el análisis desde cero, se recomienda seguir los pasos:

1. cambiar al directorio de analisis de craneos:

```
cd craneos-analisis
```

2. Crear un entorno virtual (opcional pero recomendado):

```
python -m venv .venv
```

3. Activar el entorno virtual:

```
# Windows
.venv\Scripts\activate

# Linux / macOS
source .venv/bin/activate
```

4. Instalar dependencias:

```
pip install -r requirements.txt
```

5. Ejecutar el análisis:

```
python scripts/analisis_craneos.py
```

Tras la ejecución, el script genera las principales tablas y figuras utilizadas en el informe y las almacena en `outputs/figures/`.

A.5. Consideraciones de reproducibilidad

Para asegurar consistencia en los resultados:

- Se recomienda mantener el mismo fichero de datos y su estructura (columnas y codificación de periodos).
- Si el análisis incluye procedimientos aleatorios (p.ej., bootstrap), debe fijarse una semilla (`np.random.seed(...)`) para replicabilidad exacta.
- Las rutas a figuras en el documento \LaTeX deben coincidir con las rutas reales del proyecto (por ejemplo, `outputs/figures/`).

Con estas medidas, el análisis presentado es completamente reproducible y verificable a partir de los ficheros entregados.

Anexo B: Salidas completas de los análisis

Con el objetivo de garantizar la transparencia del análisis estadístico y facilitar la verificación de los resultados presentados en el cuerpo principal del informe, en este anexo se recogen las salidas completas generadas por los distintos scripts de análisis.

Las salidas incluyen los estadísticos numéricos, valores de los contrastes y resultados intermedios que sustentan las conclusiones inferenciales expuestas. Para evitar redundancias y mejorar la legibilidad del documento principal, dichos resultados no se reproducen íntegramente en las secciones anteriores, sino que se centralizan en este anexo.

B.1. Análisis descriptivo

Las salidas completas del análisis descriptivo, que incluyen tamaños muestrales, medias, medianas, varianzas, desviaciones típicas, asimetría y curtosis para cada periodo histórico, se encuentran almacenadas en el fichero:

`outputs/descriptivos_summary.txt`

Este fichero respalda los resultados resumidos e interpretados en la Sección de Análisis Descriptivo del informe.

B.2. Comprobación de normalidad

Los resultados detallados de los tests de normalidad aplicados (Kolmogorov–Smirnov, Shapiro–Wilk y Lilliefors), incluyendo estadísticos y p-valores para cada submuestra, se encuentran disponibles en los siguientes ficheros:

`outputs/normalidad_ks.txt`

`outputs/normalidad_tests.txt`

Estas salidas constituyen la base numérica de la discusión sobre el cumplimiento del supuesto de normalidad presentada en la sección correspondiente.

B.3. Homocedasticidad

Los resultados completos del test de Levene, utilizado para evaluar la igualdad de varianzas entre los periodos predinástico temprano y tardío, se encuentran recogidos en el fichero:

`outputs/homocedasticidad_levene.txt`

B.4. Intervalos de confianza

Las salidas completas asociadas a la estimación de intervalos de confianza para la diferencia de medias, calculados para niveles de confianza del 90 %, 95 % y 99 %, se encuentran disponibles en:

`outputs/ic_diferencia_medias.txt`

B.5. Contrastes de hipótesis

Los resultados numéricos completos de los contrastes inferenciales realizados se encuentran en los siguientes ficheros:

`outputs/mann_whitney.txt`

`outputs/resultados_t_test.txt`

Estos ficheros contienen los estadísticos de contraste y p-valores que sustentan la comparación entre enfoques paramétricos (test t de Welch) y no paramétricos (Mann–Whitney U), discutida en la sección de Inferencia Estadística.

B.6. Observación final

La inclusión de este anexo permite verificar que todas las conclusiones presentadas en el informe se derivan directamente de los resultados obtenidos mediante los procedimientos estadísticos aplicados. De este modo, el análisis mantiene coherencia interna, reproducibilidad y trazabilidad completa entre los datos originales, el código ejecutado y las conclusiones finales.