

---

# Análisis estadístico del tamaño craneal de dos yacimientos arqueológicos

## Estadística

---

Bartolomé Mestre Fons

Lunes, 2 de Diciembre de 2024



Estadística  
2024-2025  
UCM - Universidad Complutense de Madrid

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Enunciados de los problemas</b>	<b>3</b>
<b>3. Ejercicio 1</b>	<b>4</b>
3.1. Apartado A . . . . .	4
3.2. Apartado B . . . . .	6
<b>4. Ejercicio 2</b>	<b>7</b>
4.1. Apartado A . . . . .	7
4.2. Apartado B . . . . .	9
<b>5. Conclusiones</b>	<b>11</b>
<b>6. Bibliografía</b>	<b>12</b>
<b>7. Apéndice</b>	<b>13</b>
7.1. Análisis Gaussiano de la Muestra 1 . . . . .	13

## 1. Introducción

El objetivo del presente trabajo es poner a prueba los conocimientos en estadística descriptiva e inferida paramétrica mediante la resolución de un conjunto de ejercicios. La base de datos corresponde a un conjunto de medidas craneales correspondientes a dos yacimientos arqueológicos. Se pretende realizar un análisis descriptivo primero sobre los datos, y a continuación averiguar el parentesco o independencia de las dos series de datos recogidos, para así obtener parámetros estadísticos que describan toda la población dentro de cierto intervalo de confianza.

## 2. Enunciados de los problemas

- **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.

1. 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 sub-muestra de los cráneos del predinástico temprano y para la sub-muestra de los del predinástico tardío. Comentar los resultados obtenidos. Estos comentarios son obligatorios.
2. Determinar si cada una de las dos sub-muestras sigue una distribución normal utilizando el test de Kolmogorov- Smirnov.

- **Ejercicio 2.**

1. 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 periodos 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?
2. 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.

**Observación:** Quiero insistir en que debéis hacer el test t para la diferencia de medias incluso si las condiciones no se cumplieran. En ese caso discutir la validez de los resultados obtenidos.



### 3. Ejercicio 1

#### 3.1. Apartado A

Se propone realizar un análisis descriptivo de los dos conjuntos de datos. Para ello, se adjuntan en la siguiente tabla los principales indicadores descriptivos de ambas muestras:

	Muestra	Valor
<b>Cuenta</b>	1	30
	2	30
<b>Media</b>	1	131.53
	2	132.47
<b>Moda</b>	1	132
	2	133
<b>Máximo</b>	1	134
	2	135
<b>Mínimo</b>	1	130
	2	131
<b>Desviación estándar</b>	1	0.82
	2	1.01
<b>Cuartil 25 %</b>	1	131.0
	2	132.0
<b>Cuartil 50 %</b>	1	131.5
	2	133.0
<b>Cuartil 75 %</b>	1	132.0
	2	133.0
<b>Rango</b>	1	4
	2	4
<b>Varianza</b>	1	0.67
	2	1.02
<b>Coefficiente de Pearson</b>	1	0.0062
	2	0.0076
<b>Coefficiente de Fisher</b>	1	0.65
	2	0.19
<b>Coefficiente de Curtosis</b>	1	1.16
	2	-0.28

Cuadro 1: Medidas descriptivas de las dos muestras.

Un primer análisis indica que ambas muestras poseen la misma cantidad de datos (30). Además, para el caso de la muestra número 2, la media es ligeramente superior a la de la muestra 1 (131.53 frente a 132.47). Si bien esto no asegura que las medias de las poblaciones mantengan dicha discrepancia, más adelante se demostrará que esto se mantiene con cierto intervalo de confianza. Por otra parte, los valores mínimos y máximos son similares (130 y 134 para la muestra 1, y 131 y 135 para la muestra 2), y las desviaciones estándares apuntan a que la muestra 2 posee una mayor dispersión de los datos (0.82 para la muestra 1 y 1.01 para la muestra 2) (y así con las varianzas). Además, realizando un análisis sobre los cuartiles, para la población 2 los cuartiles 50 % y 75 % coinciden en 133.0. En efecto, este es un indicativo de que en el 50 % de los datos se hallará el tamaño de 133.0 mm, coincidiendo efectivamente este valor con la moda. Cabe destacar que el rango en ambas muestras es el mismo (4), y el coeficiente de Pearson es cercano a 0. Esto indica que los datos no responden a una tendencia



lineal. En cuanto a como se distribuyen, en ambos casos se obtienen coeficientes de Fisher positivos (0.65 para la muestra 1 y 0.19 para la muestra 2), indicando que los valores tienden a concentrarse a la derecha del eje de simetría (mediana). Por último, el coeficiente de Curtosis para la muestra 1 es positivo (1.16), por lo que los datos se concentran alrededor de la mediana con una distribución leptocúrtica. En cambio, para la muestra 2 el coeficiente de Curtosis es negativo (-0.28), produciendo una distribución platicúrtica (valores alejados de la mediana).

A continuación, se grafican los tamaños en función de su frecuencia en un histograma, junto a su respectivo *boxplot* para analizar la distribución de los datos:

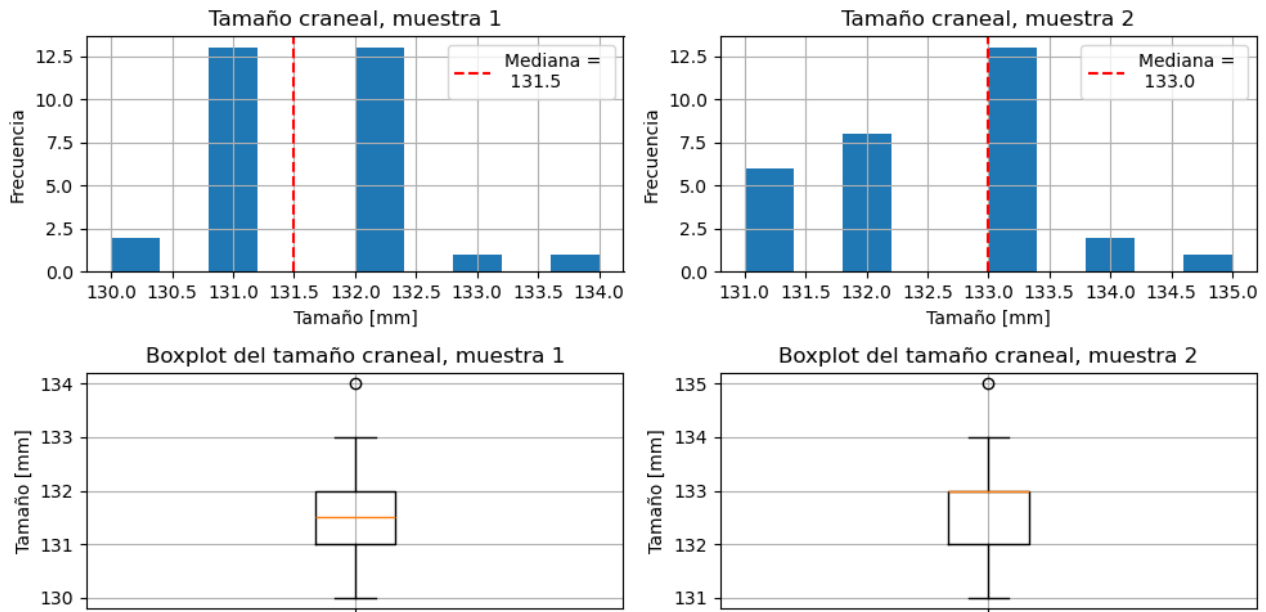


Figura 1: Histogramas y *boxplots* de ambas muestras.

Como se puede observar, en los histogramas se confirman las conclusiones anteriores: en el caso de la muestra 1, los datos se hallan más concentrados que en el caso de la muestra 2, conduciendo a un valor de Curtosis positivo versus uno negativo, respectivamente. Por otra parte, en ambos casos los datos tienden a concentrarse a la derecha de la mediana, obteniendo un valor positivo del coeficiente de Fisher. En cuanto a los *boxplots* de ambas muestras, se observa como la media (o Cuartil 50%) coincide con el Cuartil 75% como se adelantó anteriormente. También, cabe destacar la presencia de *outliers*, valores alejados de la mediana que se muestran en los *boxplots* como círculos.

### 3.2. Apartado B

A continuación, se propone implementar un test Kolmogorov-Smirnov sobre Bondad de Ajuste para determinar si las distribuciones de ambas muestras imitan una distribución normal. Por ello, las hipótesis nula y alternativa son las siguientes:

- $H_0$ : la muestra responde a una distribución normal.
- $H_1$ : la muestra no responde a una distribución normal.

El test de Kolmogorov-Smirnov sobre Bondad de Ajuste se basa en el cálculo de las diferencias entre los datos experimentales (los de las dos muestras) con los de la distribución empírica, y así también con su respectivo dato anterior. Así, se debe tomar aquella diferencia que sea mayor y compararla con el valor tabulado del coeficiente de Kolmogorov-Smirnov dado un número de grados de libertad y un intervalo de confianza deseado. Matemáticamente, se reduce a lo siguiente:

$$F_n = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & \text{si } x_i \leq x \\ 0 & \text{en otro caso} \end{cases} \quad (1)$$

donde  $F_n$  es la distribución empírica para su respectivo dato en la posición  $n$ . Entonces, siendo la distribución experimental  $F_0$  la de la muestra a considerar, las diferencias se calculan como sigue:

$$\begin{aligned} D_n^+ &= F_n(x_i) - F_0(x_i) \\ D_n^- &= F_0(x_{i-1}) - F_n(x_i) \end{aligned} \quad (2)$$

donde  $D_n^+$  es la diferencia entre el valor experimental con su respectivo empírico para el dato en la posición  $i$ , mientras que  $D_n^-$  es la diferencia entre el valor experimental anterior a  $i$  ( $i - 1$ ) respecto al valor empírico en la posición  $i$ . Entonces, el Estadístico de contraste se construye tomando el valor máximo de entre estas dos cantidades, donde cada una será un conjunto de valores (para  $n$  valores). Finalmente, dicho Estadístico debe compararse con  $D_\alpha$ , un valor tabulado dado el número de grados de libertad y el intervalo de confianza deseado ( $\alpha$ , donde si por ejemplo se desea un intervalo de confianza del 95 %, entonces  $\alpha = 1 - 0,95 = 0,05$ ) (para el presente caso, se escoge un valor de  $\alpha$  de 0.05). Si el estadístico es superior a  $D_\alpha$ , se rechaza la hipótesis  $H_0$ , y a la inversa si es menor a  $D_\alpha$ . Si bien este cálculo es tedioso para grandes cantidades de datos, Python implementa esta función con `scipy.stats.kstest`, devolviendo el valor del Estadístico y el valor P, donde si es menor que  $\alpha$  entonces se rechaza  $H_0$  y se acepta  $H_1$ . Se han obtenido los siguientes resultados:

	Estadístico	Valor P	$H_0$
Muestra 1	0.246	0.044	Rechazada (a priori)
Muestra 2	0.238	0.056	Aceptada

Cuadro 2: Resultados del Test Kolmogorov-Smirnov sobre Bondad de Ajuste.

Para la primera muestra, se obtiene un estadístico de 0.246, que comparado con el valor tabulado para 30 grados de libertad (30 datos) y  $1 - \alpha = 0,95$  se obtiene un valor superior al segundo ( $0,246 > 0,242$ , siendo este último el valor tabulado), indicando que a priori no existe suficiente evidencia para afirmar que la hipótesis nula sea verídica y que por lo tanto se pueda describir la distribución como Gaussiana. A si mismo, el valor P es inferior a  $\alpha$  ( $0,044 < 0,05$ ), confirmando la proposición anterior. Ahora bien, el poder del test Kolmogorov-Smirnov puede verse limitado por tamaños de muestra pequeños. Con  $n = 30$ , el test puede carecer de suficiente



robustez a la hora de rechazar o no la hipótesis nula. Por ello, con cierta reticencia se aceptará la hipótesis nula de que la muestra 1 puede describirse mediante una distribución normal (para más información sobre esta discusión consulte el Apéndice).

Para la muestra 2 la situación es más clara, donde el Estadístico (0.238) es inferior a su respectivo valor tabulado ( $0,238 < 0,242$ ), por lo tanto no existe en este caso evidencia suficiente para rechazar  $H_0$ , por lo que una distribución normal puede describir la muestra. A su mismo, el respectivo valor P obtenido es superior a  $\alpha$  ( $0,056 > 0,05$ ), hecho que refuerza el resultado anterior.

En conclusión, se puede verificar que con cierta reticencia y escepticismo en un 95 % de confianza la muestra 1 puede describirse mediante una distribución Gaussiana, al igual que la muestra 2 con el mismo intervalo de confianza.

## 4. Ejercicio 2

### 4.1. Apartado A

Se requiere obtener los límites para la diferencia entre las medias de ambas poblaciones dado un rango de intervalos de confianza. Para ello, deben asegurarse las siguientes tres proposiciones:

1. Las muestras deben ser independientes: asegurado por el propio enunciado del ejercicio.
2. Las muestras deben poder ser descritas mediante una distribución normal: demostrado en el ejercicio anterior.
3. Las varianzas de las poblaciones deben ser conocidas, o bien desconocidas pero iguales: se desconocen en un principio las varianzas poblacionales, y se propone a continuación determinar si son iguales o no.

Así, el problema consiste en rechazar o aceptar la siguiente hipótesis nula antes de abordar el enunciado principal:

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 && \text{Las varianzas poblacionales son iguales.} \\ H_1 : \sigma_1^2 &\neq \sigma_2^2 && \text{Las varianzas poblacionales son diferentes.} \end{aligned} \quad (3)$$

Para ello, se consulta la página 54 del Tema *Inferencia Paramétrica* referenciado en la bibliografía adjunta al final del documento. Entonces, el estadístico de contraste para este caso es el siguiente:

$$\frac{S_1^2}{S_2^2} = 0,661 \quad (4)$$

donde para evitar confusiones en la notación  $S_1$  y  $S_2$  representan las desviaciones estándares de las muestras 1 y 2, respectivamente (calculados previamente). Este Estadístico sigue una distribución F con grados de libertad  $n - 1 = 29$  y  $m - 1 = 29$ , donde  $n$  y  $m$  son los tamaños de las muestras 1 y 2, respectivamente (ambas de 30). Dado que se tienen en cuenta las dos colas de la distribución F, uno debe consultar en los datos tabulados aquellos valores de la distribución F con  $n - 1$  y  $m - 1$  grados de libertad e intervalo de confianza  $1 - \frac{\alpha}{2}$ . Se traduce a lo siguiente:

$$F_{n-1, m-1, 1-\alpha/2} < \frac{S_1^2}{S_2^2} < F_{n-1, m-1, \alpha/2} \quad (5)$$

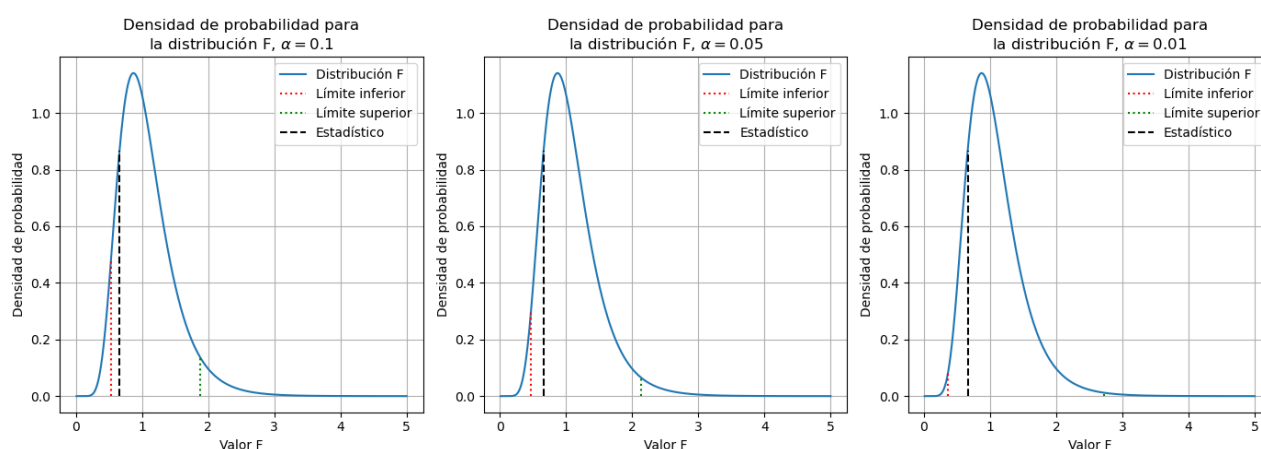


$\alpha$	$F_{n-1,m-1,1-\alpha/2}$	<b>Estadístico</b>	$F_{n-1,m-1,\alpha/2}$
0.10	0.531	0.661	1.882
0.05	0.470	0.661	2.130
0.01	0.367	0.661	2.724

Cuadro 3: Análisis de las varianzas entre las dos muestras.

Para tres valores diferentes de  $\alpha$  se ha obtenido lo siguiente:

En ambos casos el estadístico se haya comprendido entre sus límites por la distribución F, por lo que se concluye que hasta el mayor intervalo de confianza (99%) las dos varianzas poblacionales son iguales. En el siguiente gráfico se ilustran los resultados anteriores:


 Figura 2: Distribución F para tres valores diferentes de  $\alpha$  con sus respectivos Estadísticos calculados.

Con las conclusiones anteriores, ya se pueden determinar los límites de confianza para la diferencia de medias en poblaciones normales con varianzas desconocidas pero iguales. Para ello, en la página 24 del Tema *Inferencia Paramétrica* se estipulan los límites a tomar en el análisis:

$$-t_{n+m-2,\alpha/2} < \text{Estadístico} < t_{n+m-2,\alpha/2} \quad (6)$$

donde  $t_{\text{Grados de libertad},\alpha}$  es la distribución T. El Estadístico toma la siguiente forma:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(nS_1^2 + mS_2^2)(1/n + 1/m)}} \sqrt{n+m-2} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{Err} \quad (7)$$

donde  $\bar{X}$ ,  $\mu_1$  y  $\bar{Y}$ ,  $\mu_2$  son las medias de las muestras y de las poblaciones 2 y 1, respectivamente. Como se puede observar, se ha tomado la libertad de girarlas para que los intervalos de confianza sean positivos en vez de negativos, al tener la muestra 2 una media mayor a la 1. Por otra parte, *Err* representa el Error Estándar, clave a la hora de realizar los cálculos en Python. En la siguiente tabla se muestran los límites obtenidos para diferentes valores de  $\alpha$  y  $n, m = 30$ :





$\alpha$	Límite inferior	Límite superior
0.10	0.528	1.339
0.05	0.448	1.419
0.01	0.288	1.579

Cuadro 4: Intervalos de confianza para la diferencia de medias en la distribución T.

En el siguiente gráfico se muestran ilustrativamente los intervalos de confianza obtenidos:

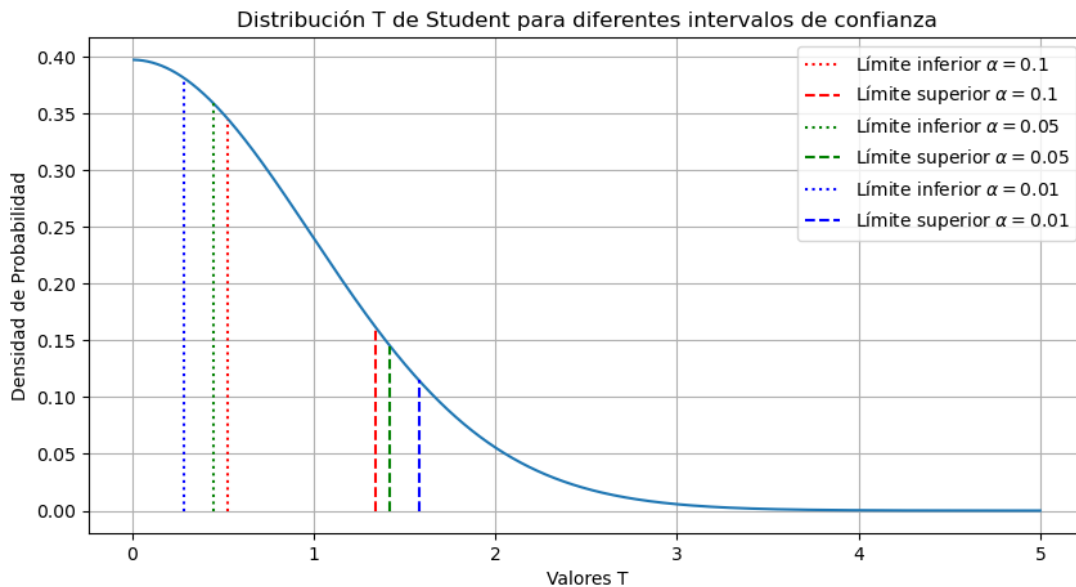


Figura 3: Distribución T para tres valores diferentes de  $\alpha$  con sus respectivos límites de confianza.

De aquí uno debe extraer la conclusión de que la diferencia entre las medias poblacionales tenderá a situarse entre los límites marcados arriba en función del valor de  $\alpha$  escogido. Manualmente, a este resultado se llega aislando  $\mu_1 - \mu_2$  de la ecuación (7) en (6) y sustituyendo numéricamente. Cabe destacar a primera vista que sea cual sea el intervalo de confianza escogido, en ningún momento el 0 se halla entre ambos límites, por lo que en ningún caso las medias podrán ser iguales, siendo la media de la población 2 siempre superior a la de la población 1.

## 4.2. Apartado B

Si bien ya se ha adelantado en el apartado anterior la imposibilidad de que las medias de ambas poblaciones puedan ser iguales, a continuación se calculará el Estadístico. Para ello, deben primero de satisfacerse las siguientes condiciones:



1. Las poblaciones deben de poder ser descritas por distribuciones normales: comprobado en el Apartado B del Ejercicio 1.
2. Las poblaciones deben de ser independientes: satisfecho por el propio enunciado del problema.
3. Las varianzas deben de ser iguales: demostrado en el Apartado A del Ejercicio 2.

Entonces, las hipótesis a considerar son las siguientes:

$$\begin{aligned} H_0 : \mu_1 = \mu_2 & \text{ Las medias poblacionales son iguales.} \\ H_1 : \mu_1 \neq \mu_2 & \text{ Las medias poblacionales son diferentes.} \end{aligned} \quad (8)$$

Así, se calcula a continuación el Estadístico enunciado en la ecuación (7), tomando en cuenta la suposición de que  $\mu_1 = \mu_2$ :

$$\text{Estadístico} = 3,849 \quad (9)$$

Ahora, se calculan los límites de la ecuación (6), que no deben confundirse con los obtenidos en el apartado A de este mismo ejercicio, ya que en dichos límites se había aislado previamente  $\mu_1 - \mu_2$  (téngase en cuenta que se trata de un análisis sobre las dos colas, por lo que a la hora de calcular la distribución T se utilizará  $\alpha/2$  en vez de  $\alpha$ ). El resultado es el siguiente:

$\alpha$	Límite inferior	Estadístico	Límite superior
0.10	-1.672	3.849	1.672
0.05	-2.002	3.849	2.002
0.01	-2.663	3.849	2.663

Cuadro 5: Análisis de las varianzas entre las dos muestras.

Gráficamente, se visualiza el Estadístico junto a los límites de la siguiente manera:



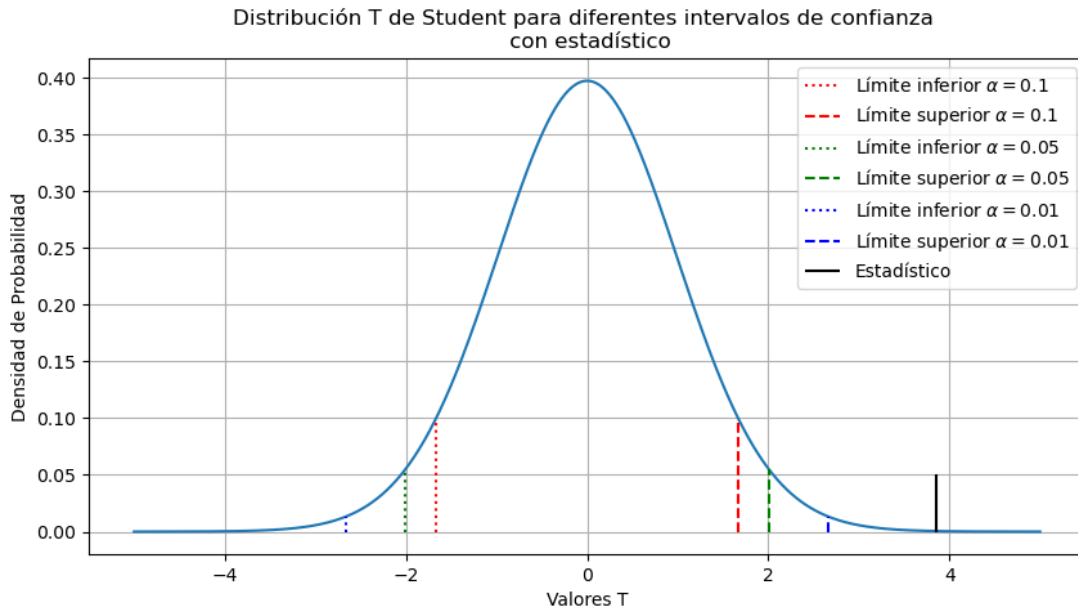


Figura 4: *Distribución T para tres valores diferentes de  $\alpha$  con sus respectivos límites de confianza y Estadístico. Se ha marcado con una línea vertical negra el Estadístico ya que cae al final de la campana, dificultando su visualización.*

Como se observa, para ningún intervalo de confianza el Estadístico cae dentro de los límites marcados, por lo que existe una fuerte evidencia para rechazar la hipótesis nula y aceptar la alternativa: las medias del tamaño craneal de ambas poblaciones son diferentes y este hecho no se debe a un factor de aleatoriedad, sino a que ambas poblaciones son estadísticamente diferentes e independientes.

Así, se confirma que los cráneos del yacimiento arqueológico 2 (predinástico tardío) poseen un tamaño superior al de los cráneos del yacimiento arqueológico 1 (predinástico temprano).

## 5. Conclusiones

Se ha llegado a las siguientes conclusiones:

- La muestra correspondiente al yacimiento datado en el predinástico tardío puede ser descrita con un 95 % de confianza como una distribución normal.
- La muestra correspondiente al yacimiento datado en el predinástico temprano requiere de un mayor examen para reconsiderar si puede describirse mediante una distribución Gaussiana. Para ello, se propone aumentar el tamaño de la muestra o bien realizar un análisis visual mediante un QQ-plot, una comparativa entre los datos experimentales con los predichos por una distribución normal. Véase el Apéndice para más información.



- Ambas poblaciones poseen varianzas iguales en un intervalo del 99 % de confianza.
- Ambas poblaciones poseen diferentes medias en el tamaño craneal.

## 6. Bibliografía

Se adjuntan a continuación los recursos bibliográficos utilizados a lo largo de la redacción y elaboración del presente documento:

- Manuel García, C. Miguel. Inferencia Paramétrica (2024). Universidad Complutense de Madrid.
- Manuel García, C. Miguel. Inferencia No Paramétrica (2024). Universidad Complutense de Madrid.



## 7. Apéndice

### 7.1. Análisis Gaussiano de la Muestra 1

Tal y como se ha adelantado en la sección Conclusiones, se propone realizar un QQ-plot para poder examinar con más minuciosidad el posible comportamiento normal de la muestra del yacimiento predinástico temprano. A continuación se adjunta el gráfico obtenido:

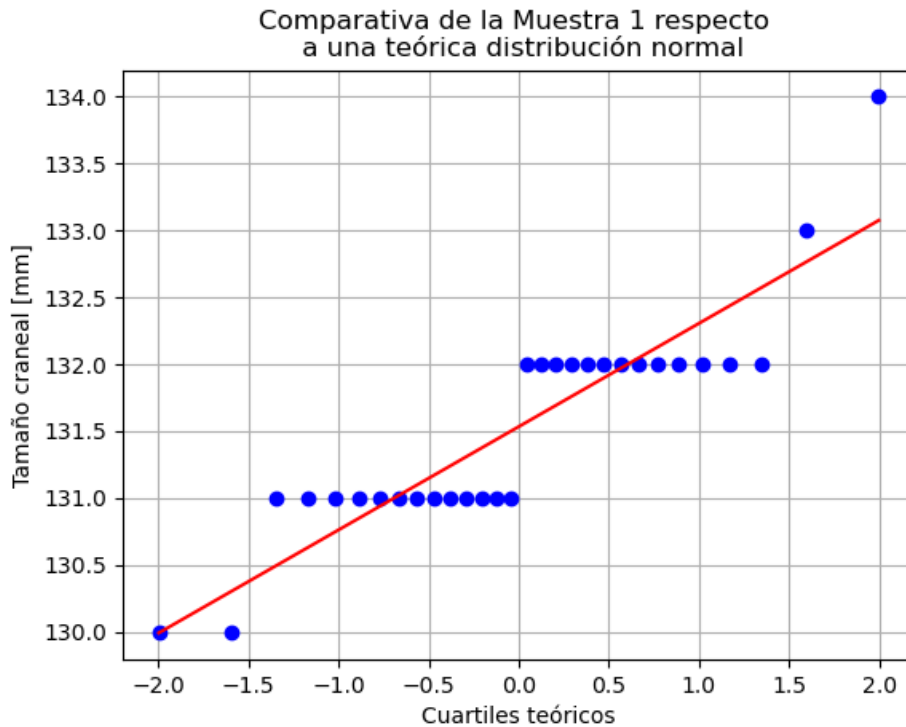


Figura 5: Datos de la Muestra 1 en comparación con los correspondientes de una teórica distribución normal. En azul se muestran los datos experimentales y en rojo la predicción gaussiana.

Se observa que para unos determinados valores del tamaño craneal hay muchos registros, y no tanto para otros valores. Así, con este análisis visual se llega a la conclusión de que, para poder decidir con mayor certeza si hay suficiente evidencia para no rechazar la hipótesis de que los datos vienen dados por una distribución normal, es necesario aumentar el tamaño de la muestra con valores más dispersos y variados, a fin tener una evidencia más sustanciosa que apoye la hipótesis nula.