INTERPRETACION DE RESULTADOS – ESTADISTICA

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

Obtendremos las Medidas de ambas muestras

MUESTRA PERIODO TEMPRANO

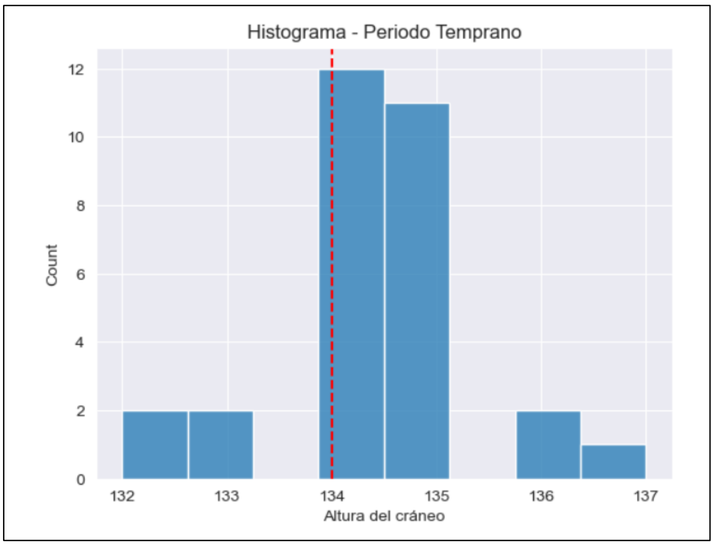
	Medida	Altura del cráneo
0	count	30.000000
1	mean	134.400000
2	std	1.069966
3	min	132.000000
4	25%	134.000000
5	50%	134.000000
6	75%	135.000000
7	max	137.000000
8	moda	134.000000
9	rango	5.000000
10	varianza	1.144828
11	CoeficientePerson	0.007961
12	CoeficienteFisher	-0.162149
13	CoeficienteCurtosis	0.540703

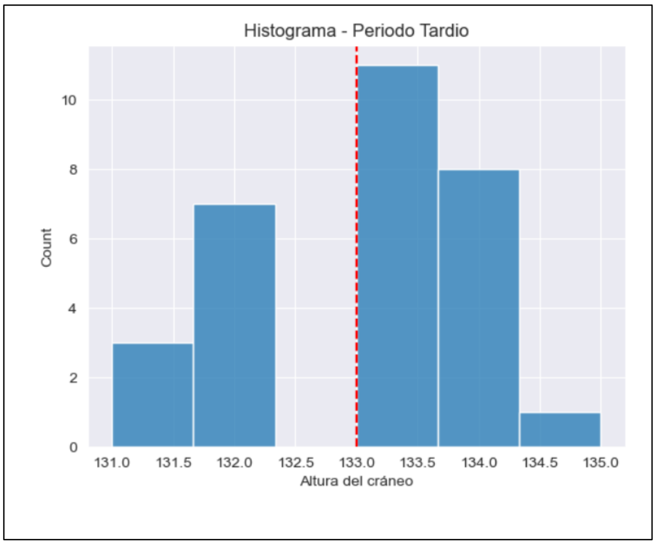
MUESTRA PERIODO TARDIO

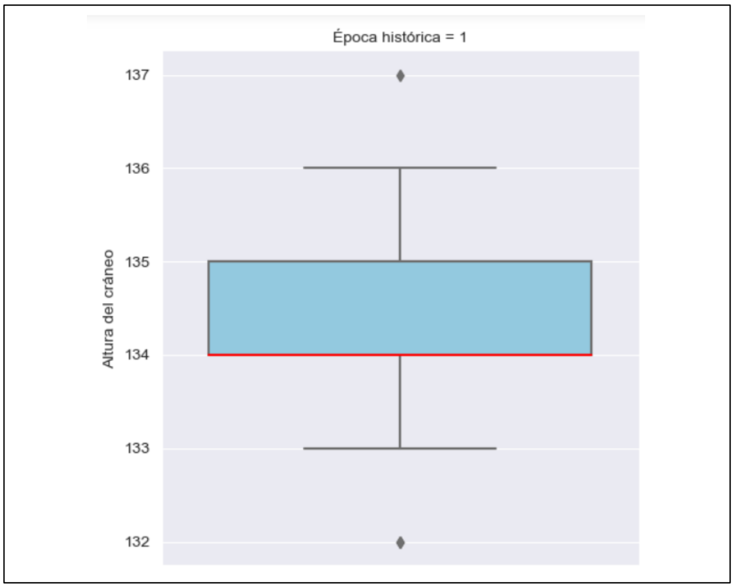
	Medida	Altura del cráneo
0	count	30.000000
1	mean	132.900000
2	std	1.028893
3	min	131.000000
4	25%	132.000000
5	50%	133.000000
6	75%	134.000000
7	max	135.000000
8	moda	133.000000
9	rango	4.000000
10	varianza	1.058621
11	CoeficientePerson	0.007742
12	CoeficienteFisher	-0.182353
13	CoeficienteCurtosis	-0.696868

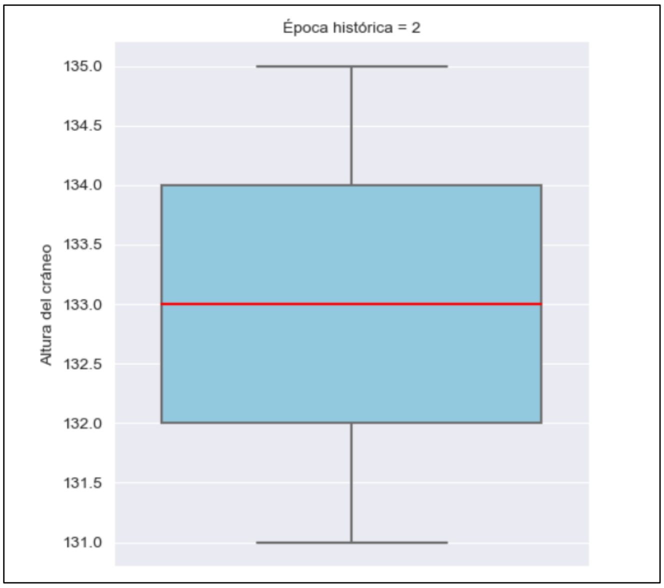
ADEMAS MOSTRAREMOS HISTOGRAMAS DE AMBAS MUESTRAS:

MUESTRA PERIODO TEMPRANO	MUESTRA PERIODO TARDIO









Bien con las Medidas calculadas en las tablas y la información adicional que nos proporcionan el histograma y el diagrama de caja. Podemos Mencionar las siguientes características de las distribuciones:

Periodo Temprano:

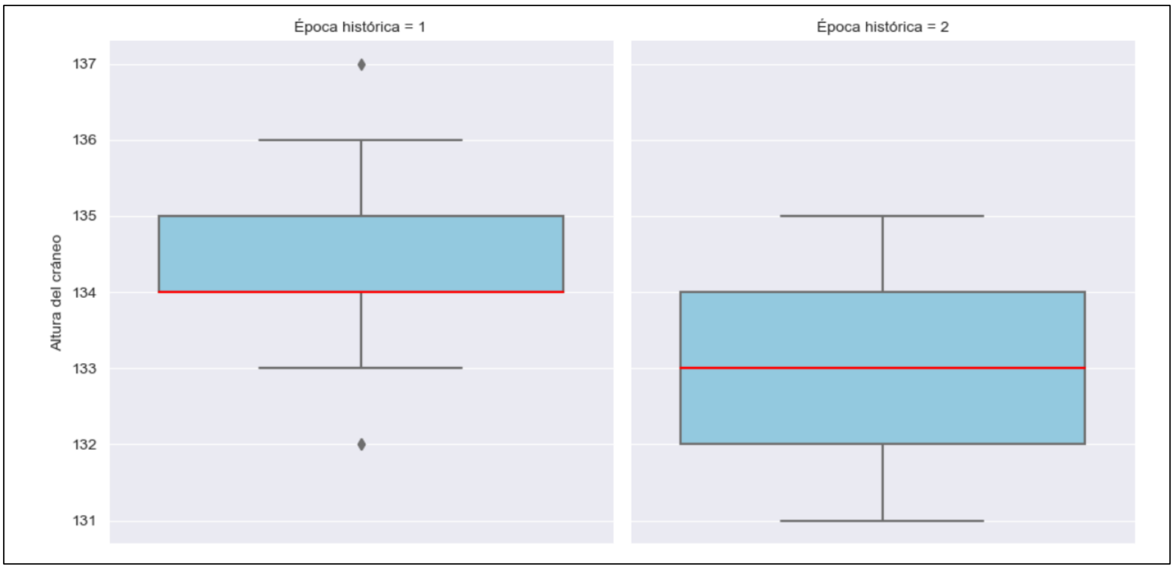
- Lo primero que podemos observar que debe llamar nuestra atención es que la Mediana (Quartil 50) y el Quartil 25 son iguales, además en el diagrama de cajas se ve como ambos valores coinciden en una línea. De esto podemos mencionar que el valor 134 se repite en el 25% de los datos.
  Dándole este 25% de repitencia del valor 134 la MODA.
- Además de este mismo diagrama de caja podemos observar que posee dos valores atípicos, muy alejado de la mediana – siendo estos valores 132 y 137

- Además, con el Coeficiente de **Asimetría con valor (-0.162149)** nos indica que los valores están mas concentrados a la izquierda del eje de simetría (mediana). cabe acotar que con la visualización del histograma pareciera que hay más valores a la derecha, sin embargo, recordemos que el valor 134 es moda, mediana y Q25
- **Coeficiente de Curtosis,** con un valor de **0.540703** lo cual nos indica una mayor acumulación de valores alrededor de la mediana y una distribución **Leptocurtica** tal como podemos visualizar en el histograma

Periodo tardío:

- Como podemos observar en el diagrama de caja del periodo tardío esta muestra es más simétrica que el periodo temprano
- Observamos que esta muestra no presenta valores atípicos
- **Coeficiente de asimetría: -0.182353** obtenemos un valor negativo para la asimetría lo que nos indica que los valores se encuentran mayormente agrupados a la izquierda. Como observamos en el histograma, siendo 133 el eje de simetría tenemos mayor cantidad de valores a la izquierda
- **Coeficiente de Curtosis: -0.6968** nos indica una distribución platicurtica la cual muestra valores menos concentrada en la mediana a comparación con la distribución normal y datos mas concentrados en las colas.

COMPARATIVA DIAGRAMAS DE CAJAS DE AMBAS MUESTRAS



Haciendo una observación rápida a la comparativa de ambas muestras sin mucho detalle ya que mas adelante entraremos en pruebas mas detalladas. El largo de los cráneos del periodo temprano es mayor a los cráneos del periodo tardío. Vemos que la mediana del temprano esta muy por encima que la mediana del periodo tardío.

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

Para la prueba de Kolmggorov tendremos:

Ho -> la distribución 1-2 sigue una distribución normal

H1 -> La distribución 1-2 no sigue una distribución normal

Para esta prueba primero normalizamos los valores (valor – Media) /DesvStd

Obteniendo los siguientes resultados:

```
Resultados de la submuestra 1:
Estadístico: 0.2185523238635185, Valor p: 0.09733554527266941
La submuestra 1 sigue una distribución normal

Resultados de la submuestra 2:
Estadístico: 0.2060393166543672, Valor p: 0.13558514721704817
La submuestra 2 sigue una distribución normal
```

COMENTARIOS DE LA PRUEBA DE KOLMOGOROV:

Muestra 1 - La muestra posee un estadístico de **0.2185** además la tabla de K-SMR para una muestra de 30 y un alfa de 0.05 el valor critico es de **0.2470** en donde cómo podemos ver el D observado es menor que el D esperado por lo que se acepta la H0
Además el pvalor es de 0.097 el cual es mayor que 0.05
Entonces concluimos: a un nivel de confianza del 95% no existe suficiente evidencia para rechazar la H0 luego la muestra 1 sigue una distribución normal

Muestra2 – La muestra posee un estadístico de **0.2060** y según la tabla K-SMR para una muestra de 30 y un alfa de 0.05 el valor critico es de **0.2470** en donde cómo podemos ver el valor D observado es menor que el D esperado por lo que se acepta la Ho
Además el pvalor es de **0.1355** el cual es mayor que 0.05
Entonces concluimos: a un nivel de confianza del 95% no existe suficiente evidencia para rechazar la H0 luego la muestra sigue una distribución normal.

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 altura 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 alta?

Para abordar este problema debemos asegurarnos de que se cumplen las siguientes condiciones:

1. Definir si las muestras son independientes – en el enunciado b se nos proporciona que asumamos independencia de muestras
2. Sabemos que las varianzas poblacionales son desconocidas
3. Demostrar si las varianzas poblacionales son iguales o diferentes

De las 3 condiciones nos falta demostrar si las varianzas poblacionales son iguales o diferentes (además esto nos ayudara en la parte b de esta pregunta)

En donde

H0: Las muestras poseen varianzas poblacionales iguales S1 = S2

H1: Las muestras poseen varianzas poblaciionales diferentes S1 != S2

Consideraciones:

- Para probar si las varianzas poblacionales son iguales o diferentes, tomaremos un intervalo de confianza del 90% -- esto un alfa/2 de 0.05
- Esta prueba es de dos colas, ya que probamos si las dsvStd son iguales o diferentes
- Además, recordemos que en la prueba de kolmorov ya determinamos que ambas muestras siguen una distribución normal.

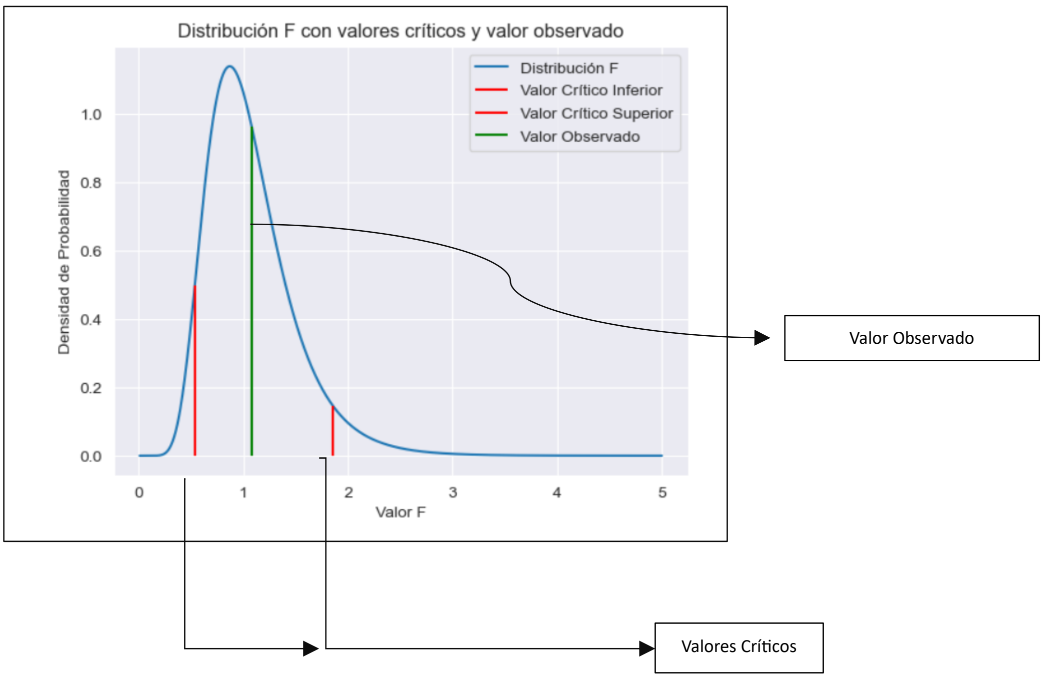
Con apoyo de Python calcularemos los valores críticos correspondientes recordando que el alfa/2 es 0.05 y ademas las muestras son de 30 dándonos grados de libertad de (29,29)

```
la zona de aprobacion de la H0 es: 0.5373999648406917 <F< 1.8608114
354760754
```

Además calculando el valor F de la división de las desv. Estándar de las muestras tenemos:

```
1.0814332247557006
```

Como observamos el valor obtenido se encuentra dentro de la zona de aceptación de la H0, gráficamente podemos mencionar que.



Del gráfico y la obtención de valores críticos y observados en la Prueba F, podemos concluir que a un 90% de confianza, no existe evidencia estadística suficiente como para rechazar la H0. Por lo tanto, se aprueba la H0 -------- con lo que las Desviaciones Estándar de ambas muestras son iguales.

Ahora, sabiendo que las condiciones bajo el cual el ejercicio opera, podemos calcular los intervalos de confianza – ya que este intervalo depende del ErrorSrandarEstimado y este varía según las condiciones.
Para la condición demuestras independientes, muestras normales y varianzas poblacionales desconocidas pero iguales tenemos de EERORSTANDARESTIMADO:

$$\frac{\sqrt{(n_1 S_X^2 + n_2 S_Y^2)[(1/n_1) + (1/n_2)]}}{\sqrt{n_1 + n_2 - 2}}$$

Con esto nos falta calcular el valor crítico T con grados de libertad (n1+n2-2) y un alfa dependiendo de que intervalo de confianza adoptemos. Con ayuda de Python tendremos el siguiente resultado:

```
los intervalos de confianza para la diferencia de medias con un 90%
de confianza es: (1.0392423859194757, 1.9607576140805243)
-------------------------------------------------------------------
-------------------------------------------
los intervalos de confianza para la diferencia de medias con un 95%
de confianza es: (0.9482336469310129, 2.051766353068987)
-------------------------------------------------------------------
-------------------------------------------
los intervalos de confianza para la diferencia de medias con un 99%
de confianza es: (0.7658743593183648, 2.2341256406816354)
```

Como observamos ya disponemos de los intervalos de confianza para la diferencia de medias (X1-X2) donde

X1 = Periodo Temprano

X2 = Periodo Tardío

Con lo que podemos concluir que al nivel de confianza de 90-95 y 99 al no incluir 0 en el intervalo de confianza, podemos mencionar que las medias poblacionales son considerablemente diferente.
Es mas al ser nuestra comparativa (X1 – X2) podemos decir que los cráneos del periodo temprano son mas alargados que el periodo 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. Observación: Quiero insistir en que debéis hacer el test t para la diferencia de medias aunque las condiciones no se cumplan. En ese caso discutir la validez de los resultados obtenidos

Abordaremos este ejercicio mencionando que para poder probar la diferencia de medias de dos poblaciones. Debemos asegurarnos de que se cumplan 3 condiciones

1. Normalidad de los datos (con la prueba de kolmogrov- smirnoy se probo que las muestrras siguen una Distribución Normal)
2. Homogeneidad de la varianza (se demostró en el ejercicio anterior que las varianzas son homogéneas con prueba F)
3. Independencia de las observaciones (el ejercicio nos exige que esta condición ya se cumple)

Entonces al satisfacer todas las condiciones para el test, plantearemos nuestras hipótesis

H0 → Las medias de ambas muestras son iguales --------- M0 = M1

H1 → Las medias de ambas muestras son diferentes ------ M0 != M1

Consideraciones:

- Como tenemos una prueba de demostrar si las medias son iguales o diferentes, esta corresponde a una prueba de dos colas
- Además, consideramos un intervalo de confianza de 95% por lo que alfa medios será de 0.025 a cada lado de las colas
- También tomamos los grados de libertad correspondientes a n1+n2-2

Con ayuda de Python, calcularemos los valores críticos de la prueba esto para la distribución T que dependerá de los grados de libertad y el alfa a cada lado de la cola que ya determinamos que es 0.025 obteniendo:

```
la zona de aceptacion para la prueba es : -2.0017174830120927 , 2.0
017174830120923
```

Calcularemos el valor observado T:

T = (Mmuestral1 – Mmuestral2) – (Xpoblacional1 – Xpoblacional2) / ErrorStandarEstimado

Como estamos probando en H0 que las medias poblacionales son iguales la diferencia representada en el calculo del T es 0 (Xpoblacional1 – Xpoblacional2 = 0)

Quedandonos el T de la siguiente manera:

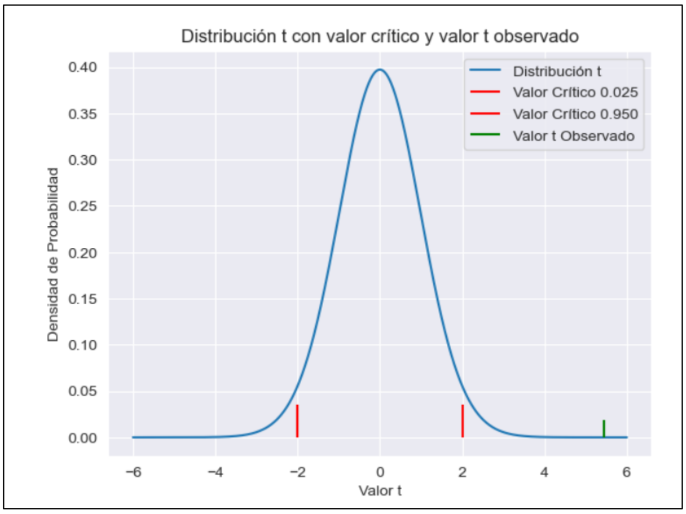
t = (Mmuestral1 - Mmuestral2) / Error Standr Estimado

recordemos que el ErrorStandarEstimado → ya lo tenemos calculado del ejercicio 2.a

nos quedaría calcular el T siendo

```
el valor critico de T es : 5.441753031545813
```

Además, podemos plantear el grafico de esta prueba siendo:



Teniendo toda esta información correspondiente podemos determinar que:

Con un intervalo de confianza del 95% se rechaza la H0 (Las medias de ambas muestras son iguales) por lo que podemos decir que las medias poblacionales son diferentes

Con esto podemos definir que la longitud de los cráneos del periodo temprano son diferentes al periodo tardío

Mas aun podemos decir que la media de la longitud de los cráneos del periodo temprano es más larga que la media de la longitud del periodo tardío, esto con apoyo del grafico

Como vemos en el grafico el valor observado esta muy a la derecha de los valores críticos indicando mayor longitud de los cráneos del periodo temprano