

Tarea Estadística - 01/04/2022

Christopher A. Rodriguez Principe

Este estudio se desarrollo con el proposito de evaluar la Anchura de Craneo sobre una muestra de la poblacion. Queremo comparar el tamaño del craneo de los individuos que vivieron durante el periodo predinastico *tardio* y hacer cmparacion con el periodo predinastico *temprano*

Presentamos los Datos de Anchura del Craneo durante periodos Predinastico Tardio(2) y Temprano(1)

```
#Dataframe de todos los datos
Datos = read_excel("~/Desktop/DS UCM /Modulo 9 /Libro1.xlsx")
as.data.frame(Datos)
```

##	Época histórica	Anchura del cráneo
## 1	1	133
## 2	1	131
## 3	1	131
## 4	1	132
## 5	1	134
## 6	1	131
## 7	1	132
## 8	1	132
## 9	1	131
## 10	1	131
## 11	1	132
## 12	1	131
## 13	1	131
## 14	1	132
## 15	1	132
## 16	1	131
## 17	1	131
## 18	1	132
## 19	1	131
## 20	1	131
## 21	1	132
## 22	1	132
## 23	1	132
## 24	1	131
## 25	1	130
## 26	1	132
## 27	1	132
## 28	1	131
## 29	1	132
## 30	1	130
## 31	2	132
## 32	2	133
## 33	2	131
## 34	2	133
## 35	2	133
## 36	2	133
## 37	2	134
## 38	2	132
## 39	2	131
## 40	2	133
## 41	2	133
## 42	2	133
## 43	2	132
## 44	2	133
## 45	2	133
## 46	2	133
## 47	2	133
## 48	2	131
## 49	2	135
## 50	2	131
## 51	2	133

```
## 52          2          132
## 53          2          132
## 54          2          132
## 55          2          131
## 56          2          134
## 57          2          131
## 58          2          133
## 59          2          132
## 60          2          132
```

Separamos los Dataframes por temporada

```
#Lista de datos Tempranos
DatosTempranos <- filter(Datos,Datos$`Época histórica`==1)

#Lista de Datos Tardio
DatosTardio <- filter(Datos,Datos$`Época histórica`==2)
```

Estadísticas de Epoca predinastica Tardio

```
#Creamo dataframe con resumen estadístico
Tardio = describeBy(DatosTardio)
Tardio <- Tardio[-c(1), ]
Tardio <- Tardio[,-c(1)]

#creamos variable pearson
Tardio$pearson <- Tardio$sd/Tardio$mean
quant = sapply(DatosTardio[-c(1)], quantile)

print(Tardio)
```

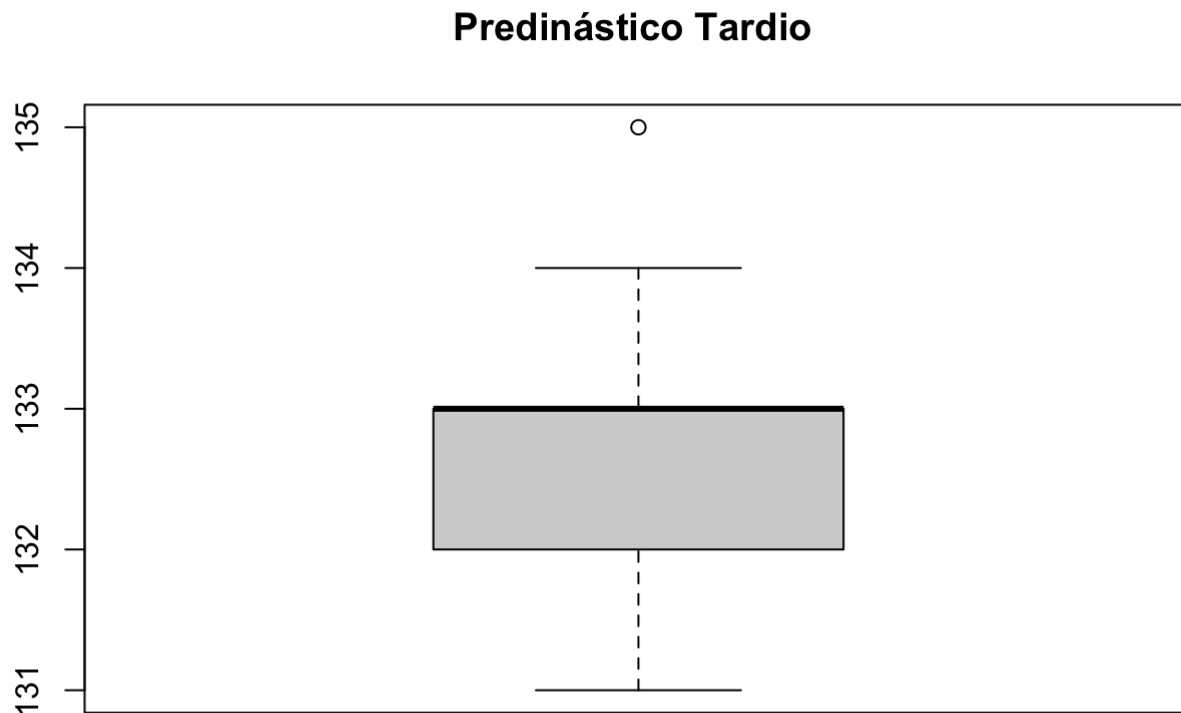
```
##          n   mean   sd median trimmed  mad min max range skew
## Anchura del cráneo 30 132.47 1.01    133  132.42 1.48 131 135     4 0.19
##          kurtosis   se pearson
## Anchura del cráneo  -0.37 0.18    0.01
```

```
print(quant)
```

```
##      Anchura del cráneo
## 0%          131
## 25%          132
## 50%          133
## 75%          133
## 100%         135
```

```
# Diagrama de caja y bigote
```

```
boxplot(DatosTardio[2], main = "Predinástico Tardio ", horizontal = FALSE)
```



Descripcion de Predinastico Tardio:

La siguiente sub-muestra de datos de cráneo del Predinastico Tardio nos presenta las siguientes medidas. La muestra contiene 30 datos de anchura del cráneo con una media, mediana y desviación media de 132.37, 133, 1.48 respectivamente. La muestra tiene un valor mínimo de anchura de cráneo de 131(mm) y anchura máxima de 135(mm) durante este periodo, esto representa un rango de solamente 4(mm) durante la duración de la época. En adición nuestra muestra presenta un coeficiente de variación pearson de ~0.01. Posee una distribución asimétrica positiva de ~0.19 demostrando que existe mayor concentración de valores a la derecha de la media que a su izquierda. Con un coeficiente de kurtosis de -0.37 presentando una distribución platycúrtica.

Presentado por la Caja de bigote. Los valores de los cuartiles son los siguientes: Q1 representa el 25% de los datos tiene anchura de 132(mm) o menos. Q3 representa el 75% de los datos tiene anchura de 133(mm) o menos. Q4 representa el ~100% de los datos tiene anchura de 135(mm) o menos.

Observamos en nuestro diagrama de caja y bigote. El tercer cuartil y la mediana coinciden con el mismo valor, esto puede ser una consecuencia de no tener variedad de anchura de los cráneos. En otras palabras los valores se repiten durante el periodo. Existe solamente una medida de nuestra muestra de valores que es atípico. La anchura de este valor atípico es de 135(mm).

Estadísticas de Epoca predinastica Temprana

```
#Creamo dataframe con resumen estadístico
Tempra = describeBy(DatosTempranos)
Tempra <- Tempra[-c(1), ]

Tempra <- Tempra[,-c(1)]

#creamos variable pearson
Tempra$pearson <- Tempra$sd/Tempra$mean
quant = sapply(DatosTempranos[-c(1)], quantile)

print(Tempra)
```

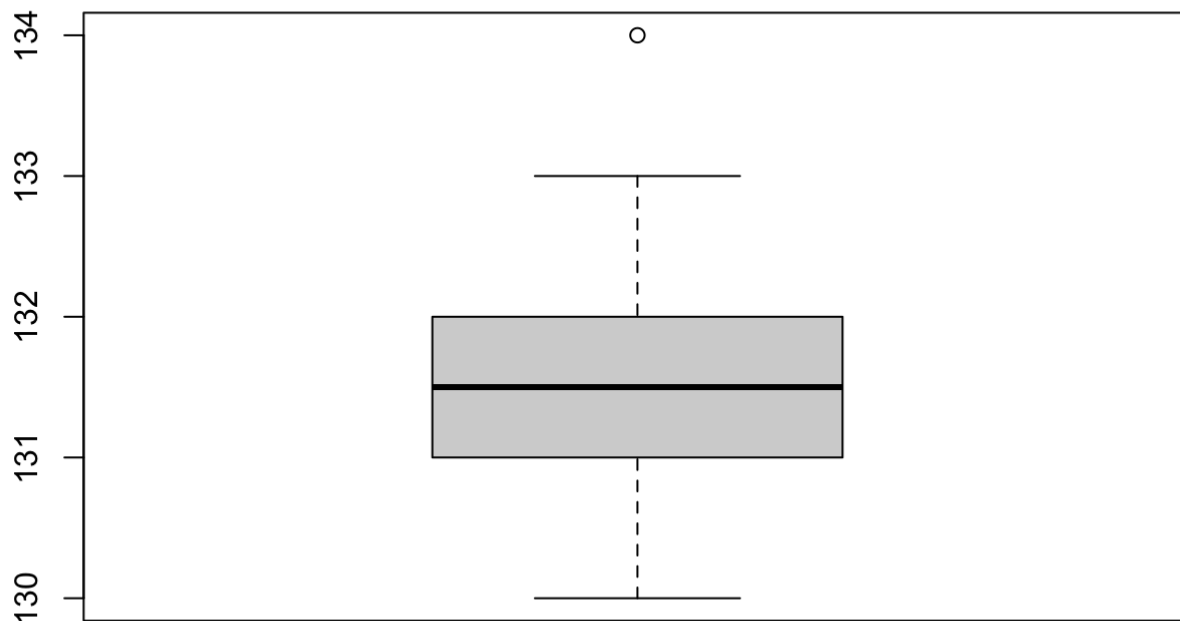
```
##              n   mean   sd median trimmed  mad min max range skew
## Anchura del cráneo 30 131.53 0.82  131.5   131.5 0.74 130 134     4 0.62
##              kurtosis    se pearson
## Anchura del cráneo    1.02 0.15     0.01
```

```
print(quant)
```

```
##      Anchura del cráneo
## 0%                130.0
## 25%                131.0
## 50%                131.5
## 75%                132.0
## 100%               134.0
```

```
#Desplegamos el diagrama de caja y bigote
boxplot(DatosTempranos[2], main = "Predinástico Temprano ", horizontal = FALSE)
```

Predinástico Temprano



Descripcion de Predinastico Temprana:

La siguiente sub-muestra de datos de cráneo del Predinastico Tardio nos presenta las siguientes medidas. La muestra contiene 30 datos de anchura del cráneo con una media, mediana y desviación media de 131.53, 131.5, 0.74 respectivamente. La muestra tiene un valor mínimo de anchura de cráneo de 130(mm) y anchura máxima de 134(mm) durante este periodo, esto representa un rango de solamente 4(mm) durante la duración de la época. En adición nuestra muestra presenta un coeficiente de variación pearson de ~0.01. Posee una distribución asimétrica positiva de ~0.62 demostrando que existe mayor concentración de valores a la derecha de la media que a su izquierda. Con un coeficiente de kurtosis de 1.02 presentando una distribución leptocúrtica.

Presentado por la Caja de bigote. Los valores de los cuartiles son los siguientes: Q1 representa el 25% de los datos tiene anchura de 131(mm) o menos. Q3 representa el 75% de los datos tiene anchura de 132(mm) o menos. Q4 representa el ~100% de los datos tiene anchura de 134(mm) o menos.

Nuestra submuestra presenta solamente un valor atípico de 134(mm) durante el periodo predinástico tardío.

Test Lilliefors

```
#Test Temprano
lillie.test(DatosTardio$`Anchura del cráneo`)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  DatosTardio$`Anchura del cráneo`
## D = 0.23496, p-value = 0.0001938
```

```
#Test Tardio
lillie.test(DatosTempranos$`Anchura del cráneo`)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  DatosTempranos$`Anchura del cráneo`
## D = 0.24246, p-value = 9.677e-05
```

Con nuestras dos muestras continuas de Datos Predinastico Temprano y Tardio intentamos describir una prueba de distribución normal. La primera opción para esta prueba debería ser el KS.test pero dado que tenemos muy pocos datos y existen varios elementos en la sub-muestra que son repetidos hemos considerado el *Lilliefors test* para estudiar la distribución.

En nuestra prueba hemos establecido dos hipótesis (H_0 y H_1):

H_0 : La muestra proviene de una distribución normal.

H_1 : La muestra NO proviene de una distribución normal.

Para *DatosTardio* p-value = 0.0001938. Dado p-value = 0.0001938 < 0.05 concluimos que la anchura del cráneo para el periodo predinástico tardío NO sigue distribución normal. Se rechaza H_0 .

Para *DatosTempranos* p-value = 9.677e-05 Dado p-value = 9.677e-05 < 0.05 concluimos que la anchura del cráneo para el periodo predinástico temprano NO sigue distribución normal. Se rechaza H_0 .

Condiciones de un T test para una muestra independiente:

Explico qué condiciones se deben cumplir para contrastar la hipótesis de que ambas medias son iguales.

Independencia - Las observaciones tienen que ser independientes unas de las otras. Para ello el muestreo debe ser aleatorio y el tamaño de la muestra inferior al 10% de la población.

Resultado = *Verdadero*

Normalidad- Las poblaciones que se comparan tienen que distribuirse de forma normal.

Hemos demostrado en el apartado anterior en el ejercicio 1.b que nuestros datos NO poseen distribución normal. Resultado = *Falso*

Igualdad de varianzas- a varianzas de ambas poblaciones comparadas debe de ser igual.

Resultado = *Verdadero*

Comprobamos con el Levene- Test si las varianzas de ambos sub-muestras se asemejan. H_0 : Las varianzas de las dos muestras son iguales. H_1 : Las varianzas de las dos muestras NO son iguales.

```
#comprobamos si las varianzas son similares.
leveneTest(DatosTempranos$`Anchura del cráneo`, DatosTardio$`Anchura del cráneo`)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  4  0.4388 0.7793
##      25
```

$p_value = 0.7793 > 0.05$. NO se puede rechazar la hipótesis de que las varianzas de ambas muestras son iguales. *Resultado* = las varianzas son iguales. NO rechazamos H_0 .

Test sobre igualdad de las medias

H_0 : Las medias de las dos muestras son iguales.

H_1 : Las medias de las dos muestras NO son iguales.

A pesar de que no se cumplen las tres condiciones para el T-test, lo implementamos para observar como son las medias muestradas.

T-test

```
#Calculamos los intervalos con un nivel de confianza de 0.90, 0.95 y 0.99.
```

```
t.test(x=DatosTempranos$`Anchura del cráneo`, y=DatosTardio$`Anchura del cráneo`, conf.level = 0.99)
```

```
##
## Welch Two Sample t-test
##
## data: DatosTempranos$`Anchura del cráneo` and DatosTardio$`Anchura del cráneo`
## t = -3.9354, df = 55.675, p-value = 0.0002329
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -1.5658551 -0.3008116
## sample estimates:
## mean of x mean of y
## 131.5333 132.4667
```

```
t.test(x=DatosTempranos$`Anchura del cráneo`, y=DatosTardio$`Anchura del cráneo`, conf.level = 0.95)
```



```
##
## Welch Two Sample t-test
##
## data: DatosTempranos$`Anchura del cráneo` and DatosTardio$`Anchura del cráneo`
## t = -3.9354, df = 55.675, p-value = 0.0002329
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.4084846 -0.4581821
## sample estimates:
## mean of x mean of y
## 131.5333 132.4667
```

```
t.test(x=DatosTempranos$`Anchura del cráneo`,y=DatosTardio$`Anchura del cráneo`,conf.level = 0.90)
```

```
##
## Welch Two Sample t-test
##
## data: DatosTempranos$`Anchura del cráneo` and DatosTardio$`Anchura del cráneo`
## t = -3.9354, df = 55.675, p-value = 0.0002329
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
## -1.3300289 -0.5366378
## sample estimates:
## mean of x mean of y
## 131.5333 132.4667
```

Ahora evaluamos el p-value nuestro t-test para los datos predinásticos tardíos y tempranos. Obtenemos un p-value = 0.0002329. Ahora evaluamos si se rechaza nuestra hipótesis de si las medias son iguales:

P-value(0.0002329) es menor que 0.05.

Por lo tanto concluimos que hay evidencia en contra de que las medias de las dos muestras son iguales. Se rechaza la hipótesis nula.

En adición, observamos que los intervalos de confianza son negativos:

99 percent confidence interval: -1.5658551 -0.3008116

Si hacemos un cálculo de $\text{mean}(\text{Temprano}) - \text{mean}(\text{Tardio}) = -0.93$ Esa resta nos demuestra que las medias de la muestra es más predominante sobre la época predinástica Tardía.

Conclusión

Dado que el intervalo de confianza es menor que 0 podemos deducir que la anchura de los cráneos es mayor durante la época predinástica tardía que en la época predinástica temprana.

$(\text{Anchura época tardía}) > (\text{Anchura época temprana})$